

Universidade Federal de São Paulo – UNIFESP
Escola Paulista de Política, Economia e Negócios - EPPEN
Departamento de Ciências Atuariais

Bárbara Ivânia Rangel Prazeres

DEFINIÇÃO DE GRUPOS DE RISCO DE ROUBO OU FURTO, PERDA PARCIAL
OU TOTAL DE VEÍCULOS VIA K-MEANS

Osasco

Outubro de 2020

Bárbara Ivânia Rangel Prazeres

DEFINIÇÃO DE GRUPOS DE RISCO DE ROUBO OU FURTO, PERDA PARCIAL
OU TOTAL DE VEÍCULOS VIA *K-MEANS*

Trabalho de Conclusão de Curso apresentado como
requisito parcial à obtenção do título de Bacharel,
Curso de Ciências Atuariais, Escola Paulista de
Política, Economia e Negócios, Universidade Federal
de São Paulo.

Orientador: Prof. Dr. Raphael de Oliveira Garcia

Osasco

Outubro de 2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha catalográfica elaborada pela Biblioteca Unifesp Osasco
e Departamento de Tecnologia da Informação Unifesp Osasco,
com os dados fornecidos pelo(a) autor(a)

com os dados PRAZERES, Bárbara Ivânia Rangel de
fornecidos Definição de grupos de risco de roubo de automóveis via
pelo(a) *kmeans* / Bárbara Ivânia Rangel de Prazeres. - 2020.
autor(a)com 60 f. :il.
os dados
fornecidos Trabalho de conclusão de curso (Ciências Atuariais)
pelo(a) Universidade Federal de São Paulo - Escola Paulista de
Política, Economia e Negócios, Osasco, 2020.
autor(a)P921d Orientador: Raphael de Oliveira Garcia.

1. Ciência atuarial. 2. Automóveis - Roubo. 3. Gênero.
4. Seguros. I. Garcia, Raphael de Oliveira, II. TCC
Unifesp/EPPEN. III. Título.

CDD: 368.01

Bárbara Ivânia Rangel Prazeres

Trabalho de conclusão de curso aprovada,
em cumprimento às exigências legais como
requisito parcial para à obtenção do título de
Bacharel em Ciências Atuariais.

Aprovada em: ____/____/____

BANCA EXAMINADORA

Prof. Dr. Raphael de Oliveira Garcia – Orientador

Profa. Dra. Graciele Paraguaia Silveira - UFSCar/Sorocaba

EPÍGRAFE

“A quem tenho eu nos céus senão a ti? E na terra, nada mais desejo além de estar junto a ti. O meu corpo e o meu coração poderão fraquejar, mas Deus é a força do meu coração e a minha herança para sempre”.

Salmos 73:25,26

AGRADECIMENTOS

Agradeço primeiramente a Deus, o meu Senhor, Redentor e Pai, por me conceder sabedoria e me proporcionar o aprendizado nessa jornada da graduação. Agradeço ao Orientador Professor Raphael de Oliveira Garcia pela dedicação, paciência e generosidade, a banca examinadora. A meu esposo Jurandir Prazeres Filho, minha inspiração; pelo seu imenso amor, compreensão, incentivo, apoio, paciência, e pelo homem maravilhoso que é, como poderei te agradecer? Não tem dinheiro no mundo que pague..., mas Deus o Todo Poderoso, esse sim te abençoe de forma abundante e transbordante. Você é um dos motivos para que eu pudesse chegar até aqui.

אוהב אותך כל כך אהובי.

RESUMO

Os índices de roubo na cidade de São Paulo, entre 2018 e 2019, indicaram uma queda no número de furtos e roubos de veículos de acordo com o Ministério da Justiça (2020). Segundo a revista Veja, em 2019, a Zona Leste foi a região paulistana que reuniu seis dos dez bairros campeões de furtos de veículo. Em geral, o risco de roubo ou furto do veículo está relacionado a localização (bairro ou distrito) e consequentemente ao preço do seguro. Todavia será que isso apenas ocorre para roubo ou furto? E quanto a perda parcial ou perda total? Seria possível agrupar os distritos segundo a frequência de sinistros ou somente agrupá-los de acordo com a proximidade geográfica explicaria o risco? Com o objetivo de estudar estas frequências de sinistros considerando a geolocalização foram utilizadas as bases de dados disponibilizadas pela Superintendência de Seguros Privados (SUSEP) através do Sistema de Estatísticas de Automóveis da SUSEP (Autoseg). As bases de dados analisadas continham 60 milhões de registros (CEP, ano modelo, marca, categoria do veículo, exposição, número de sinistros de roubo ou furto, número de sinistros de perda total ou parcial). Estas bases de dados foram sumarizadas por distrito, um total de 96, onde as frequências de sinistros de roubo ou furto e perda total ou parcial foram calculadas no período de 2016 a 2018. Estas frequências foram plotadas no mapa e fora aplicado o algoritmo k-means considerando a frequência de roubo ou furto obtendo 12 clusters e ao repetir o procedimento para frequência de perda parcial ou total obteve-se igualmente 12 clusters. Em geral, distritos circunvizinhos apresentaram frequência de sinistro similares, o que torna razoável agrupá-los segundo a proximidade geográfica, no entanto, ao utilizar o k-means, distritos de zonas distintas foram inseridos dentro do mesmo grupo, o que evidencia que outros aspectos influenciam no risco e devem ser considerados.

Palavras-chave: Roubo de automóvel, Roubo de veículo, Seguro, Perda Total, Perda Parcial, Valor do Seguro

ABSTRACT

Theft rates in the city of São Paulo, between 2018 and 2019, indicated a drop in the number of vehicle thefts and thefts according to the Ministry of Justice (2020). According to *Veja* magazine, in 2019, the East Zone was the São Paulo region that brought together six of the ten neighborhoods that were champions of vehicle thefts. In general, the risk of theft or theft of the vehicle is related to the location (neighborhood or district) and consequently the insurance price. However, does this only occur for theft or theft? What about partial loss or total loss? Is it possible to group districts according to the frequency of claims or just grouping them according to geographical proximity already explains the risk? In order to study these claim frequencies considering the geolocation, the databases made available by the Superintendency of Private Insurance (SUSEP) through the SUSEP Automobile Statistics System (Autoseg) were used. The databases analyzed contained 60 million records (CEP, model year, make, vehicle category, exposure, number of theft or theft claims, number of claims for total or partial loss). These databases were summarized by district, a total of 96, where the frequency of theft or theft claims and total or partial loss were calculated in the period from 2016 to 2018. These frequencies were plotted on the map and the k- means considering the frequency of robbery or theft obtaining 12 clusters and when repeating the procedure for frequency of partial or total loss, 12 clusters were also obtained. In general, surrounding districts had similar frequency of claims, which makes it reasonable to group them according to geographical proximity, however, when using the k-means districts from different zones were grouped, which shows that other aspects influence the risk and should be considered.

Keywords: Auto theft, Vehicle theft, Insurance, Total Loss, Partial Loss, Insurance Amount

LISTA DE ILUSTRAÇÕES

Gráfico 1 - Estatística de roubo ou furto de 2014 a 2109.....	16
Quadro 1 - Descrição das tabelas contidas no bando de dados da Autoseg- Sistema de Estatística de Automóveis da SUSEP.....	21
Quadro 2 - Descrição das variáveis utilizadas para o estudo no bando de dados da Autoseg- Sistema de Estatística de Automóveis da SUSEP.....	22
Figura 1 - Representação de Clusters.....	24
Figura 2 - Centróides C_{11} , C_{21}	26
Figura 3 - Centróides C_{12} , C_{22}	27
Figura 4 - Como funciona o K-means.....	28
Mapa 1 - Mapa dos distritos do Estado de São Paulo.....	32
Mapa 2 - Exposição 2016.....	34
Mapa 3 - Distritos dos Municípios de São Paulo com a frequência de roubo ou furto em 2016.....	35
Mapa 4 - Distritos dos Municípios de São Paulo com a frequência de PP ou PT em 2016.....	36
Mapa 5 - Exposição 2017.....	37
Mapa 6 - Distritos dos Municípios de São Paulo com a frequência de roubo ou furto em 2017.....	38
Mapa 7 - Distritos dos Municípios de São Paulo com a frequência PP ou PT em 2017.....	39
Mapa 8 - Exposição 2018.....	40
Mapa 9 - Distritos dos Municípios de São Paulo com a frequência de roubo ou furto em 2018.....	41
Mapa 10 -Distritos dos Municípios de São Paulo com a frequência PP ou PT em 2018.....	42
Mapa 11 – Análise <i>k-means</i> frequência de roubo ou furto em 2016.....	44
Mapa 12 – Análise <i>k-means</i> frequência de roubo ou furto em 2017.....	45
Mapa 13 – Análise <i>k-means</i> frequência de roubo ou furto em 2018.....	46
Mapa 14 – Análise <i>k-means</i> frequência de PP e PT em 2016.....	48
Mapa 15 – Análise <i>k-means</i> frequência de PP e PT em 2017.....	49
Mapa 16 – Análise <i>k-means</i> frequência de PP e PT em 2018.....	50
Tabela 1 – Elementos A,B,C e D.....	26

Tabela 2 – Frequência de sinistro por ano.....	32
Tabela 3 - Quantidade de distrito e frequência de roubo ou furto por <i>cluster</i> por ano.....	43
Tabela 4 - Quantidade de distrito e frequência de perda parcial ou total por <i>cluster</i> por ano.....	47
Quadro 3 - Distritos e <i>clusters</i> para roubo ou furto 2016.....	52
Quadro 4 - Distritos e <i>clusters</i> para roubo ou furto 2017.....	53
Quadro 5 - Distritos e <i>clusters</i> para roubo ou furto 2018.....	54
Quadro 6 - Distritos e <i>clusters</i> perda parcial ou total 2016.....	55
Quadro 7- Distritos e <i>clusters</i> perda parcial ou total 2017.....	56
Quadro 8 - Distritos e <i>clusters</i> perda parcial ou total 2018.....	57

LISTA DE ABREVIATURAS E SIGLAS

CASCO – É o termo designado para a estrutura do carro, desde sua lataria, elétrica, mecânica e tudo que compõem o modelo dentro de suas especificações originais.

CNT – Confederação Nacional dos Transportes

CNSP - Conselho Nacional de Seguros Privados

CTB – Código de Trânsito Brasileiro

DETRANSP – Departamento de Trânsito de São Paulo

FECAP - Fundação Escola de Comércio Álvares Penteado

FUNENSEG – Fundação Escola Nacional de Seguros

IBGE - Instituto Brasileiro de Geografia e Estatística

IPEA – Instituto de Pesquisa Econômicas Aplicadas

PIB – Produto Interno Bruto

PRF – Polícia Rodoviária Federal

PP – Perda Parcial

PT – Perda Total

R – Software de análise de dados estatísticos

SP – São Paulo

SSP- SP - Secretaria de Segurança Pública de São Paulo

SIG – Sistema de Informação Geográfico

SUSEP - Superintendência de Seguros Privados

Sumário

1.Introdução.....	13
2. Caracterização da área de estudo.....	15
2.1. Conceitos básicos.....	15
2.2. O cenário dos furtos e roubos de automóveis na capital de São Paulo.....	18
2.3 Causas de sinistros com perda parcial e total	19
2.4 Preços praticados pelas seguradoras no mercado de automóvel em São Paulo.....	19
3. Metodologia.....	21
3.1. Dados	21
3.2. Método k-means	24
4. Análises e Resultados	31
5. Considerações Finais	59
6. Referências Bibliográficas.....	61
7. Anexos.....	64

1. Introdução

O Brasil convive com diferenças regionais bastante acentuadas sobretudo no índice de roubo de veículos. Enquanto alguns distritos apresentam índices de roubo e furto de veículos baixos, outros distritos, por sua vez, possuem índices de roubo alarmantes (G1 Globo, 2017). Além disso, as políticas de segurança pública têm falhado no combate ao crime de roubo de veículos.

Apesar de algumas regiões apresentarem melhora no índice de roubo nos últimos dois anos, o Brasil teve, em 2018, um recorde negativo nas estatísticas criminais: um mês com um maior número de carros roubados desde 2003 (Dia, 2018). Embora deva-se considerar os problemas relacionados às informações como no Estado do Rio de Janeiro onde há má qualidade de preenchimento dos boletins de ocorrência nas delegacias, além do alto número de erros e inconsistências nas poucas informações registradas (pessoas nascidas em 1800, motociclistas com 5 anos de idade, etc.). Em alguns casos há ausência de informações sobre rua e/ou bairro, logradouros que não correspondem aos bairros, confusão do bairro de ocorrência com o bairro sede da circunscrição policial. Estes problemas dificultam um pouco o estudo uma vez que acarretam em perda de informações (MUSMECI; CONCEIÇÃO, 2018).

Em um cenário mais atual, segundo informação do portal Agência Brasil, O Ministério da Justiça divulgou um balanço sobre crimes cometidos em todo o país. O comparativo, entre os primeiros semestres de 2018 e 2019, mostra uma queda no número de furtos e roubos de veículos. Com relação a furto e roubo de veículos, a queda ficou em 10,9% (de 40.527 para 36.123 ocorrências), e 28,3% (43.448 para 31.161 registros) do ano anterior.

Um aumento no roubo de veículos, assim como a falha na recuperação destes veículos, afeta diretamente o preço das apólices de seguros uma vez que os preços são agravados em função deste índice e, consequentemente, há menos apólices emitidas e renovação de cobertura. Em muitos casos, em áreas onde o risco é maior, o valor do seguro dobra. Quando um veículo é roubado e desmanchado, a seguradora tem de quitar o valor integral do veículo e isso tem um impacto no preço final do seguro (O Globo, 2018). Por outro lado, uma diminuição no índice de roubo reflete em preços mais baratos. Uma vez que cada seguradora utiliza suas próprias regiões de risco para definir os preços, um agrupamento de regiões considerando frequência de roubo e furto de veículos, perfil do segurado e dados

dos veículos do mercado pode auxiliar o processo de precificação de uma seguradora. Neste trabalho considerou-se que as regiões são distritos definidos de acordo com o CEP de circulação. A base de dados é fornecida pela SUSEP com informações das frequências de roubo de 2016 a 2018 e encontra-se disponível na internet através do site *Autoseg* – Sistema de Estatísticas de Automóveis da Susep.

Uma das maneiras de construir estes agrupamentos ou *clusters* de risco de roubo de veículos é via a utilização de técnicas de estatística multivariada - *clustering*. A estatística multivariada consiste em um conjunto de técnicas estatísticas utilizadas quando as variáveis são medidas simultaneamente. Neste projeto de pesquisa, especificamente, a ideia é utilizar um método chamado *k-means* cuja finalidade é agregar observações com base nos atributos disponíveis que pode ser frequência de sinistro de roubo e furto, perfil do segurado, dados do veículo, etc.

O objetivo do *k-means* é classificar as observações em um pequeno número de grupos mutuamente excludentes de acordo com similaridades existentes entre eles. O método classifica as observações segundo aquilo que cada uma tem de similar em relação a outra.

Segundo Reis (2011), os cálculos envolvidos na estatística multivariada são facilmente e rapidamente feitas num computador com um *software* apropriado instalado. A questão é escolher o método adequado ao tipo de dados e usá-lo corretamente, saber interpretar os resultados e retirar deles as conclusões corretas.

Este trabalho tem como objetivo aplicar o algoritmo *k-means* para criar *clusters* de distritos segundo as frequências de sinistros de roubo ou furto e as frequências de perda parcial ou total e entender como o risco para estes dois tipos de perda difere entre os distritos da cidade de São Paulo.

No capítulo 2, é apresentada a caracterização da área de estudo através de definições e cenários de roubo ou furto e perda parcial ou total no Brasil. No terceiro capítulo é apresentada a metodologia utilizada no trabalho com os dados utilizados da SUSEP, além da descrição da técnica *k-means* e do uso do QGIS para construção dos mapas. No quarto capítulo são apresentadas as análises, resultados e interpretação com o uso de tabelas, quadros e mapas. Por último, no quinto capítulo, as considerações finais são apresentadas além de sugestão de trabalhos futuros.

2. Caracterização da área de estudo

Neste trabalho optou-se por estudar a frequência de sinistro de furto e/ou roubo, perda total e/ou perda parcial de veículos nos distritos da cidade de São Paulo uma vez que os preços, numa mesma seguradora, para o mesmo perfil, variam bastante entre os bairros do município. Este capítulo destinou-se para a caracterização da pesquisa. Na Seção 2.1 apresenta-se os conceitos básicos envolvidos na contratação de seguro, dados estatísticos de roubo e/ou furto nos subdistritos da cidade de São Paulo, além de conceitos do software k-means e QGIS. Na Seção 2.2 um breve cenário dos furtos e roubos de veículos na capital de São Paulo e na Seção 2.3 as causas e caracterização dos sinistros com perda parcial e total.

2.1 Conceitos básicos

De acordo com o Código Penal - Decreto-Lei nº 2.848, de 7 de dezembro de 1940, Furto é subtrair, para si ou para outrem, coisa alheia móvel (Art. 155). E Roubo é Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência a pessoa, ou depois de havê-la, por qualquer meio, reduzido à impossibilidade de resistência (Art. 157).

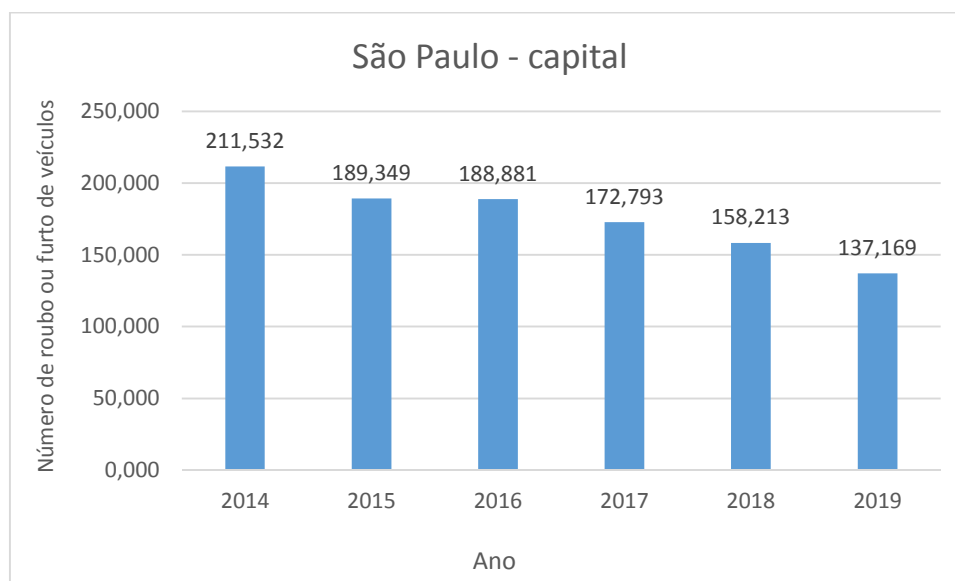
O sinistro é qualquer tipo de incidente que causa dano ou prejuízo, em que a seguradora precise ser acionada. Após avaliar os danos, a seguradora os classifica como perda parcial ou total.

De acordo com o Minuto Seguro, perda parcial é quando o veículo sofre danos em que os custos de reparação não atingem 75% do valor indicado na Tabela Fipe que é a principal referência para pessoas que querem comprar ou vender veículos usados ou seminovos, e tem por objetivo auxiliar o consumidor a realizar, o cálculo de depreciação do veículo, além de ser utilizada para definir o IPVA e o cálculo do seguro. E a Perda Total é quando o veículo sofrer algum acidente e os valores de reparo superarem 75% do valor do carro.

Segundo informações fornecidas por uma reportagem do site de notícias R7 (2020), os roubos e furtos de veículos no Estado de São Paulo ultrapassaram a marca de 1 milhão de casos entre 2014 e 2019, de acordo com boletins de ocorrência policiais registrados pela SSP-SP (Secretaria de Estado da Segurança Pública de São Paulo).

De acordo com os dados da reportagem mencionada, o ano de 2014 foi o que registrou o maior número de casos (211.532). Nos anos seguintes, houve decréscimo gradual da quantidade de ocorrências: 189.349 (2015), 188.881 (2016), 172.793 (2017), 158.213 (2018) e 137.169 (2019) ver Gráfico 1.

Gráfico 1. Estatísticas de roubo ou furto de 2014 a 2019



Fonte: Secretaria de Segurança Pública do Estado de São Paulo (2020, Apud R7, 2020)

Este crime tem se tornado um ato bem lucrativo para as organizações criminosas, pois de acordo com os dados do Ministério da Infraestrutura do Governo Federal, atualizados em fevereiro de 2020, a frota do Estado de São Paulo é 30.131.254 veículos. Na capital paulista, há 8.583.039 veículos licenciados, e as quadrilhas tem se especializado cada vez mais neste tipo de ação (R7, 2020).

Ainda em uma reportagem da revista Veja de 2019, a Zona Leste é citada como a região que reúne seis dos dez bairros campeões de furtos de veículo na cidade de São Paulo. Do total, a maioria dos casos (8.836) ocorre de manhã, enquanto a madrugada é o período do dia menos comum (1.951) para furtos do tipo.

Porém observou-se que a maioria das ocorrências foram nas vias públicas (23.422 casos), embora a capital também registre casos em estacionamentos pagos e de comércios e serviços, dentre outros (VEJASP, 2019).

A reportagem ainda aponta que um dos motivos para haver menos roubos nas regiões mais próximas do centro é a maior presença de policiamento e segurança.

Segundo IBGE a divisão territorial brasileira é estruturada em Macrorregiões, Unidades da Federação, Mesorregiões, Microrregiões e Municípios, além de suas subdivisões internas, os distritos e subdistritos ou regiões administrativas. Os subdistritos são unidades geográficas que dividem o território do distrito ou do município.

Portanto, para fins de estudo do presente trabalho, tem-se que ‘Sinistro’ é um acidente que causa danos e/ou prejuízos a um bem segurado, especificado na apólice do seguro. E independe da forma do acontecimento, se súbita, involuntária ou imprevista. Por isso todos os danos causados pelas situações cobertas na apólice deverão ser indenizados. O sinistro pode ser parcial ou total, a depender do dano causado. A perda parcial acontece quando o veículo pode ser reparado. E o sinistro integral (perda total) quando o carro não pode ser recuperado, seja por motivo de roubo ou colisão (MeuPortoSeguro, 2015). Todos esses dados, compreendem o estudo dos subdistritos de São Paulo que atualmente é composto por 96, irão compor a pesquisa sobre a influência da localização na determinação dos preços das apólices de seguros, praticados nas regiões de São Paulo pelo mercado segurador de automóvel e ajudar a entender os fatores que contribuem para essa diferença.

O risco está presente em qualquer atividade desenvolvida, o que implica em determinadas tomadas de decisão. Neste trabalho o risco tem um papel considerável levando em conta a quantidade de sinistros ocorridos no Estado de São Paulo. Temos várias definições de risco segundo (2007, Apud Spink 2001: 1279). Tem-se que risco é “a possibilidade de ocorrência de eventos vindouros, em um momento histórico onde o futuro passava a ser pensado como passível de controle”. Temos ainda que um ‘risco’ é uma probabilidade de que uma pessoa sofrerá um dano devido a uma ameaça em particular. E mais precisamente para o mercado segurador, de acordo com a SUSEP, é um evento incerto ou de data incerta que independe da vontade das partes contratantes e cuja ocorrência dará direito à indenização descrita na apólice.

Temos ainda o risco coletivo que segundo Ferreira (2002), é a distribuição de sinistro de uma carteira como um todo, sem se preocupar com as características dos sinistros produzidos por cada apólice, que seria o risco individual. No risco coletivo é necessário conhecer a distribuição do valor de cada sinistro, independente da apólice à qual o sinistro pertence, e conhecer a distribuição do número total de sinistros produzidos em uma carteira.

Logo o risco é incerto, não se pode determinar seu tempo, nem extensão, no entanto é sempre possível através de medidas evitar ou atenuar.

A *clusters analysis*, é uma técnica de análise multivariada com propósito de reunir objetos, baseando-se em suas próprias características. Ela classifica os objetos por similaridades entre grupos pré-selecionados. Estes grupos selecionados devem ter alto grau de homogeneidade interna (*within-cluster*) e alta heterogeneidade externa (*between-cluster*). Logo, seu objetivo é encontrar e separar objetos em grupos similares (CORRAR, et.al, 2007).

Baseado no problema proposto pelo estudo, a análise de *clusters* será usada para colocar os objetos semelhantes em *clusters*, e para isto será utilizado o algoritmo *k-means* que trabalha para encontrar *k clusters* diferentes no conjunto de dados. O centro de cada *cluster* será chamado centróide e terá a média dos valores neste *cluster*, o objetivo do algoritmo é encontrar o centróide mais próximo do objeto determinado e atribuir o ponto encontrado a esse *cluster*, ou seja, é encontrar similaridades entre os dados e agrupá-los conforme o número de *cluster* passado pelo argumento *k*. (HONDA, 2017).

O *k-means* é um algoritmo do tipo não supervisionado, ou seja, que não trabalha com dados rotulados. O processo do *k-means* é composto por quatro etapas: inicialização, atribuição ao *cluster*, movimentação de centróides e a otimização do *k-médias* que é o mesmo que *k-means*.

Também será utilizado o software livre QGIS, ele é uma ferramenta de sistema de informação geográfica que possibilita a aplicação de técnicas de geoprocessamento (SEMAD, 2019).

2.2 O cenário dos furtos e roubos de automóveis na capital de São Paulo

Em 2020 tivemos um cenário bem atípico não só em São Paulo como no mundo, com a Pandemia da COVID-19, evidenciando que quanto menos se estiver nas ruas, menor é a exposição ao risco. Estar exposto ao roubo e ao furto de automóveis está ligado a diversos fatores, principalmente ao de sair de casa. Segundo dados da Secretaria de Segurança Pública de SP (2020, Apub Revista Cobertura 2020), os índices de roubo e furto de veículos caíram 21,25% e 20,62%, respectivamente, no primeiro quadrimestre de 2020, comparados com o mesmo período de 2019, no Estado de São Paulo.

O Boletim mostra que em abril houve uma redução de 45,53% no número de roubo de veículos, e 49,29% nas ocorrências de furto de veículos, na comparação com abril de 2019. Os dados ainda revelam uma mudança de comportamento da criminalidade, pois a maioria das ocorrências no primeiro quadrimestre ocorreram durante o período noturno (48,28%). Na contramão de 2019 que no mesmo período registro maior ocorrência pela manhã (27,9%), e os índices do período noturno representavam apenas 25,2%.

2.3 Causas de sinistros com perda parcial e total

Podemos citar algumas causas de ocorrência de sinistro como: roubo ou furto, colisão, danos a terceiros, causas naturais enfim.

Em se tratando de perda total ela ocorre nas situações em que o dano atinge o percentual de 75% de não recuperação do veículo e isso inclui, furto, roubo (se este não for encontrado), colisão, alagamento e incêndio. Já a perda parcial como dito anteriormente é quando o dano não atinge o percentual de 75%, por exemplo, pequenas mosas.

Um tipo de perda parcial seria a quebra dos vidros do carro, e fatores que podem contribuir para este sinistro seria: vias com baixa quilometragem, infraestrutura precária das vias, falta de segurança, eventos climáticos, entre outras.

Para a perda total temos, eventos climáticos, roubo, acidentes de trânsito, ou seja, quando o custo do conserto do veículo dado pelo sinistro ultrapassa o percentual do valor do veículo estipulado na apólice.

Foi possível constatar que não existem muitos dados percentuais sobre ocorrência de sinistros PP (perda parcial) e PT (perda total), o que corrobora para a importância desta pesquisa.

2.4 Preços praticados pelas seguradoras no mercado de automóvel em São Paulo

Dentro de um mercado tão competitivo como das seguradoras, escolher um seguro que tenha atributos de qualidades e preços ainda assim em São Paulo não é algo tão fácil, pois apesar da concorrência os preços praticados têm um alto custo.

Os preços aplicados pelas seguradoras, tem como base o risco do bem segurado sofrer um sinistro, como: furto, roubo, perda total, perda parcial, pequenas avarias.

De acordo com uma matéria da revista Quatro Rodas em 2019, um dos motivos que encarece o seguro de automóvel, são as coberturas e os serviços inclusos na apólice, no entanto o segurador tem o poder de escolher o plano que melhor atende suas necessidades. Mas mesmo assim, os preços continuam não lhe sendo favorável em termos econômicos.

Já as seguradoras por sua vez, criam pacotes com configurações diversas para atender a demanda do cliente, buscando uma solução para melhorar o preço e ter mais poder de venda. Para as seguradoras o que encarece as apólices de seguros são serviços e itens como: reboque e emergência, carro reserva, franquias, idade, sexo, estado civil, motorista adicional, o índice de roubo dos veículos também é levado em consideração, pois existem carros para os quais a taxa de roubo é muito maior do que a de outros (MINUTOSEGUROS, 2020).

Segundo dados da SSP-SP em 2019, foram 46.517 mil carros roubados e 90.642 mil carros furtados na capital de São Paulo. Até junho de 2020 o índice é de 16.222 e 33.101 respectivamente.

Em reportagem da Veja São Paulo, em 2018 a média de veículos furtados diariamente na capital foi de 109. Dentre os quais o destaque em número de ocorrência fica para 14 bairros: 1º São Mateus: 855/ 2º Sapopemba: 718/ 3º Ipiranga: 683 / 4º Vila Mariana: 649/ 5º São Lucas: 643/ 6º Tatuapé: 641/ 7º Vila Prudente: 632/ 8º Penha: 587/ 9º Tucuruvi: 564/ 10º Perdizes: 540/ 11º Itaquera.

Após a caracterização da área de estudo, no Capítulo 3 é apresentada a metodologia empregada neste trabalho.

3. Metodologia

A metodologia envolve a apresentação da base de dados, definição do k-means e suas principais características.

3.1 Dados

A base de dados utilizada neste estudo é de propriedade pública, esses dados são coletados, informados e gerenciados pela SUSEP (Superintendência de Seguros Privados). A SUSEP é um órgão responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro. Autarquia vinculada ao Ministério da Economia, e foi criada pelo Decreto-lei nº 73, de 21 de novembro de 1966.

Segundo informações da própria SUSEP, suas atribuições são:

- Fiscalizar a constituição, organização, funcionamento e operação das Sociedades Seguradoras, de Capitalização, Entidades de Previdência Privada Aberta e Resseguradores, na qualidade de executora da política traçada pelo CNSP;
- Atuar no sentido de proteger a captação de poupança popular que se efetua através das operações de seguro, previdência privada aberta, de capitalização e resseguro;
- Zelar pela defesa dos interesses dos consumidores dos mercados supervisionados;
- Promover o aperfeiçoamento das instituições e dos instrumentos operacionais a eles vinculados, com vistas à maior eficiência do Sistema Nacional de Seguros Privados e do Sistema Nacional de Capitalização;
- Promover a estabilidade dos mercados sob sua jurisdição, assegurando sua expansão e o funcionamento das entidades que neles operem;
- Zelar pela liquidez e solvência das sociedades que integram o mercado;
- Disciplinar e acompanhar os investimentos daquelas entidades, em especial os efetuados em bens garantidores de provisões técnicas;
- Cumprir e fazer cumprir as deliberações do CNSP e exercer as atividades que por este forem delegadas;
- Prover os serviços de Secretaria Executiva do CNSP.

Os dados podem ser encontrados na SUSEP através do endereço eletrônico: <http://www2.susep.gov.br/menuestatistica/Autoseg/principal.aspx>, pelo sistema AUTOSEG (Sistema de Estatística de Automóveis da SUSEP), que permite realizar consultas on-line referentes a dados estatísticos do Seguro de Automóveis. As informações nele contida são obtidas a partir dos arquivos enviados semestralmente pelas companhias seguradoras, os quais incluem dados referentes a apólices vigentes e sinistros ocorridos no período de análise.

Os dados utilizados referem-se aos roubos, furtos, perda parcial e total em São Paulo, entre os anos 2016 e 2018, e as variáveis utilizadas neste estudo se encontram nos Quadros 1 e 2:

Quadro 1 - Descrição das tabelas contidas no banco de dados da Autoseg- Sistema de Estatística de Automóveis da SUSEP.	
Tabelas	Descrição
Tabela Arq_Casco_Comp	Contém dados de exposição, prêmios, sinistros e importância segurada para a cobertura CASCO, classificados pela chave Categoria Tarifária/Região/Modelo/Ano/Sexo/Faixa Etária
arq_casco3_comp	Contém dados de exposição, prêmios e sinistros para a cobertura CASCO, classificados pela chave Categoria Tarifária/CEP/Modelo/Ano
arq_casco4_comp	Contém dados de exposição, prêmios e sinistros para a cobertura CASCO, classificados pela chave Categoria Tarifária/Cidade/Modelo/Ano
(continua)	

Quadro 1 - Descrição das tabelas contidas no banco de dados da Autoseg- Sistema de Estatística de Automóveis da SUSEP.	
(conclusão)	
Tabelas Auxiliares	
auto2_vei	Contém código e descrição de cada modelo de veículo, além do código do grupo a que pertence
auto2_grupo	Código e descrição dos grupos de modelos
auto_cat	Código de descrição de categorias tarifárias
auto _ cau	Código e descrição de causas de sinistros
auto_CEP	Correlaciona o CEP com cidades e regiões de circulação
auto_cob	Código e descrição de coberturas
auto_idade	Código e descrição de faixas etárias
auto_reg	Código e descrição de regiões de circulação
auto_sexo	Código e descrição de sexo (masculino, feminino, jurídico)

Fonte: SUSEP / AUTOSEG, 2020. Adaptada pela autora.

Quadro 2 - Descrição das variáveis utilizadas para o estudo no banco de dados da Autoseg- Sistema de Estatística de Automóveis da SUSEP.	
Variáveis	Descrição
EXPOSICAO	Quantidade de veículos expostos
PREMIO	Soma dos valores de prêmios, ponderados pela exposição de cada apólice
FREQ_SIN1	Quantidade de sinistros da cobertura roubo/furto
INDENIZ1	Total de indenizações de sinistros da cobertura roubo/furto
FREQ_SIN23	Quantidade de sinistros da cobertura colisão parcial
INDENIZ23	Total de indenizações de sinistros da cobertura colisão parcial

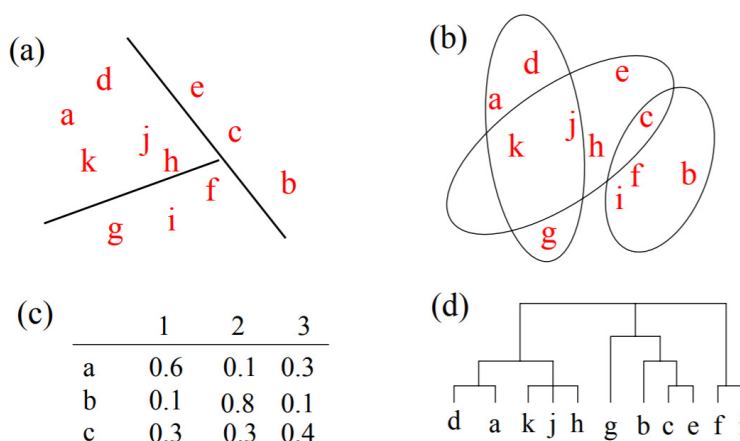
Fonte: SUSEP / AUTOSEG, 2020. Adaptada pela autora.

3.2 Método k-means

Este trabalho foi embasado na análise de dados via estatística descritiva, que é uma das áreas da estatística que cuida em explicar os dados de um determinado conjunto, e cujo objetivo é condensar uma série de valores de mesma natureza, proporcionando uma visão comum desses valores, em que se organiza e expõe os dados por meio de gráficos, tabelas e medidas descritivas (GUEDES et al, 2018).

O *clustering* é um conjunto de técnicas que agrupa observações predefinidas, como a criação de subgrupos dentro de um determinado conjunto de dados. E é desta forma que este algoritmo trabalha, ele agrupa as observações em classes distintas, de modo que, as observações na mesma classe devem ser semelhantes entre si, enquanto as observações dentro de grupos diferentes são classificadas como dissimilares entre si, ver Figura 1 (OLIVEIRA, 2019).

Figura 1. Representação de *Clusters*



Fonte: USP - <http://dfm.ffclrp.usp.br/~augusto>

Para Oliveira (2019), o *k-means clustering* é um método usado no aprendizado de máquina, nele se extrai inferências de dados com respostas não rotuladas. Logo, de posse de um conjunto de observações o algoritmo tentará agrupar as observações em um número pré-especificado de *k clusters* distintos e não sobrepostos. Este método é realizado pelo agrupamento não-hierárquico é uma técnica usada quando se deseja formar *k* grupos de itens ou objetos. O *k-means* também é considerado como um algoritmo de mineração de dados

não supervisionado, pois fornece uma classificação de informações de acordo com os próprios dados.

Segundo Nunes (2016), o *k-means* busca encontrar o *clustering* que minimiza a soma de quadrado do resíduo, convergindo para uma solução local em vez de um *clustering* global. O algoritmo distribui aleatoriamente os pontos do conjunto D em k *clusters* e calcula os centróides através da média dos pontos do cluster C_i (centróide inicial). Em seguida o método busca a atribuição de *clusters* e a atualização dos centróides (o centróide e a base de cada grupo de k objetos). Na atribuição de *clusters*, cada ponto $x_j \in D$ (a função D mede a semelhança entre o ponto x_j e o centróide z_i , no *k-means*) é associado ao *cluster* que tem o centróide z_i (centróide do *cluster* i) mais próximo do ponto, isto é, x_j é atribuído ao *cluster* C_j^* quando:

$$C_j^* = \operatorname{argmin}_{i=1,2,\dots,k} \{\|x_j - z_i\|_2^2\}$$

onde *argmin* é o argumento do mínimo, ou seja, é a otimização do valor mínimo da função $x_j - z_i$, e $\|\cdot\|_2$ representa a norma-2 euclidiana. Norma é a forma de calcular a distância entre os elementos de um dado conjunto.

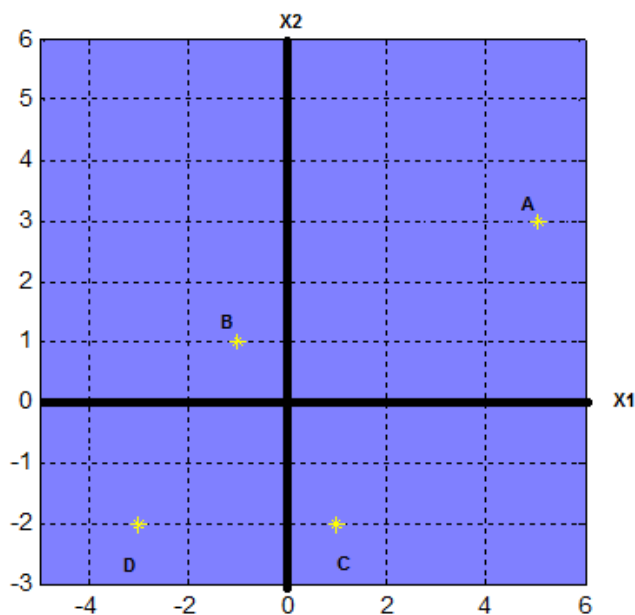
O algoritmo utiliza a distância euclidiana para calcular o quão ‘longe’ uma ocorrência está da outra. E desta forma utiliza ligações mais simples como (vizinhos mais próximos), juntando-se dois grupos com menor distância ou maior similaridade.

Um exemplo ilustrado por Vânia Bogorny (2020) auxilia na explanação sobre a execução do algoritmo *k-means*. Considere um subconjunto do plano cartesiano formado pelos seguintes elementos: $A = (5,3)$, $B = (-1,1)$, $C = (1,-2)$ e $D = (-3,-2)$. Ver Tabela 1 e a disposição desses elementos no plano cartesiano ver Figura 2. Além disso, considere que o objetivo seja agrupar os elementos A , B , C e D em dois grupos, $k=2$. O primeiro passo do método é determinar os centróides iniciais (escolhidos em geral aleatoriamente). Supondo que foram sorteados os itens A e B , logo $C_{11} = (5,3)$ e $C_{21} = (-1,1)$.

Tabela 1. Elementos A, B, C e D

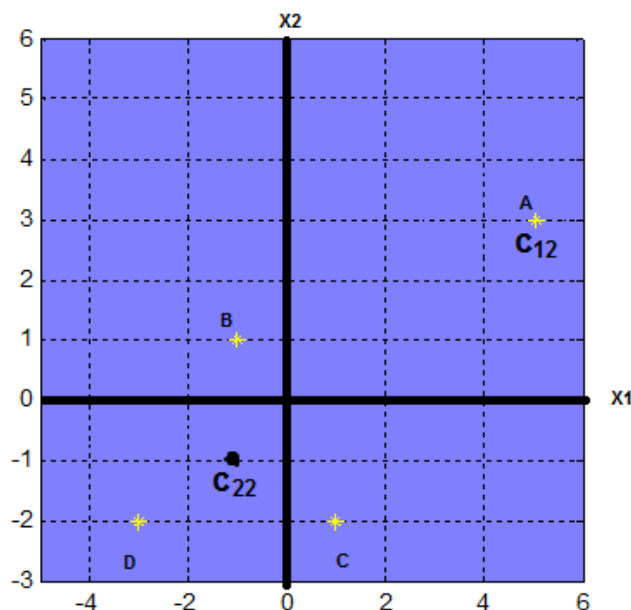
Elementos	Coordenadas	
	x1	x2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

O segundo passo envolve a definição dos *clusters* iniciais (Figura 3) através do cálculo da distância de cada ponto aos centróides.

Figura 2. Centróides C_{11} , C_{21} 

Fonte : <http://www.inf.ufsc.br/~vania/teaching/INE5644/clustering1.ppt>

Como $k = 2$ estão definidos dois *clusters*, $C_1 = \{A\}$, $C_2 = \{B, C, D\}$, pois C e D estão mais próximos do centróide C_{21} (item B) do que do centróide C_{11} (item A). O terceiro passo envolve o cálculo de novos centróides, $C_{12} = (5, 3)$, $C_{22} = ((x_B + x_C + x_D)/3, (y_B + y_C + y_D)/3) = ((-1 + 1 - 3)/3, (1 - 2 - 2)/3) = (-1, -1)$. No passo 4, define-se novos clusters. Neste passo, nota-se que como os *clusters* não se alteram, os centróides serão os mesmos e então o algoritmo para.

Figura 3. Centr ides C_{12} , C_{22} 

Fonte : <http://www.inf.ufsc.br/~vania/teaching/INE5644/clustering1.ppt>

Conforme a Figura 3, os clusters finais s o $C_1 = \{A\}$ e $C_2 = \{B, C, D\}$ uma vez que A est  mais pr ximo de C_{12} do que C_{22} e B, C e D est o mais pr ximos de C_{22} do que de C_{12} .

O *k-means* atualiza os centr ides atrav s da m dia de todos os pontos que se encontram no *cluster* C_i , iterativamente at  alcan ar um m nimo local, isto  , ele converge se os centr ides n  mudarem ap s uma itera  o.

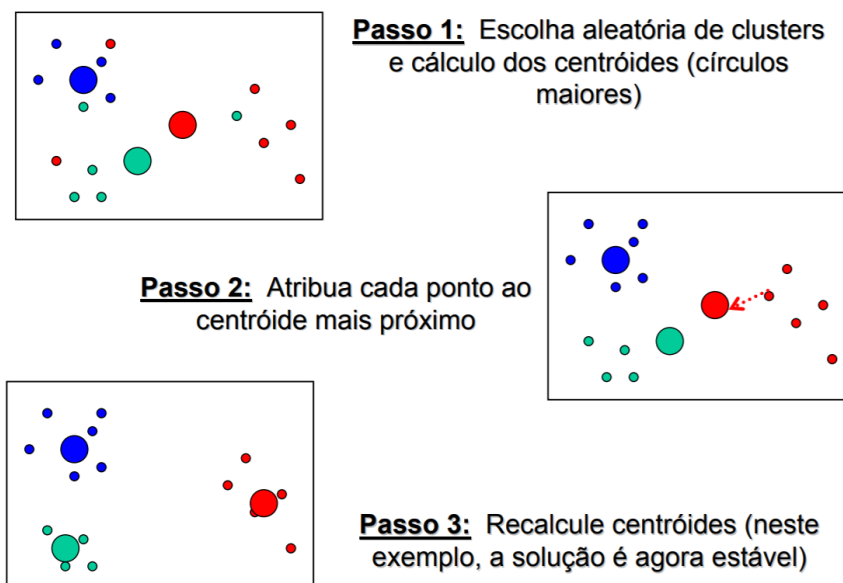
O algoritmo possui uma condi  o de parada como no exemplo:

$$\sum_{i=1}^k \|z_i^t - z_i^{t-1}\|_2^2 \leq \epsilon, \text{ onde } \epsilon > 0$$

  o limite de converg ncia, t   a itera  o corrente e z_i^t   o centr ide do *cluster* C_i na itera  o t , ϵ   um n mero suficientemente pequeno.

Em resumo o algoritmo gera k (ou menos) clusters da seguinte forma: escolhe-se k *clusters*, calcula-se o centr ide para cada *cluster*, repete-se, atribui-se a cada ponto o centr ide mais pr ximo e recalcula-se o centr ide para cada *cluster* at  a estabilidade.

Figura 4 – Como funciona o k-means



Fonte: <http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Clustering.pdf>

O *software* QGIS também foi utilizado neste trabalho. Este *software* foi escrito nas linguagens Python e C++, é um software livre (Windows, Mac, Linux entre outros) que suporta vários formatos vetoriais, *rasters* e bases de dados e outras funcionalidades. Ele tem como objetivo realizar análise de dados georreferenciados, ou seja, um processo de reconhecimento de coordenadas geográficas, e a partir de sua utilização constrói-se mapas ou imagens.

Temos que Geoprocessamento são todas as tecnologias utilizadas para aquisição, processamento, armazenamento, manutenção, interpretação e/ou análise de dados e informações georreferenciadas (SANTOS, 2018 Apud DOMINGUES, 2007).

Um SIG segundo Santos, (2018) Apud Teixeira, (1995), é um "conjunto de programas (softwares), equipamentos (hardware), metodologias (procedimentos), dados e pessoas (usuário), perfeitamente integrados, de forma a tornar possível a coleta, o armazenamento, o processamento e a análise de dados georreferenciados, bem como a produção de informação derivada de sua aplicação".

O levantamento dos dados geográficos é feito por meio do processo de Georreferenciamento, ou seja, na aplicação deste trabalho é o levantamento topográfico para definir as características referentes à incidência de furtos, roubos, PP ou PT e quilometragem das vias, dimensão e localização dos subdistritos.

No Brasil um ambiente SIG, geralmente, adota como referencial cartográfico o sistema de coordenadas geográficas geodésico SIRGAS 2000. Geodésico vem de Geodésia a ciência que tem por objetivo determinar a forma e as dimensões da Terra e os parâmetros definidores do campo da gravidade (SANTOS, 2018, Apud GEMAEL, 1994).

Pelo processo de georreferenciamento é definido quais são os limites, características e confrontação de cada subdistrito. Essas definições são feitas por meio da descrição das coordenadas dos vértices limitantes referenciadas no Sistema Geodésico Brasileiro.

Nesta etapa as bases foram carregadas no R e aplicado o filtro para cidade de SP, após esse processo realizou-se a sumarização por CEP, ou seja, soma da exposição, número de sinistro de roubo ou furto, número de sinistro de perda total ou parcial por CEP via função *aggregate*.

Então um *merge* desta base foi realizado com outra base composta por CEP, bairro e distrito da cidade de SP para adicionar o distrito na base através da chave CEP, esta etapa foi realizada no Excel. E novamente sumarizou-se a base, desta vez, por distrito, ou seja, somou-se a exposição, número de sinistro de roubo ou furto, número de sinistro de perda total ou parcial por distrito. Assim cada distrito tem suas respectivas frequências de sinistros. Esta etapa também foi realizada no Excel.

Após obter a base final sumarizada por distrito, o software QGIS foi utilizado para análise georreferenciada. O arquivo de mapa *shapefile* utilizado foi baixado do site de mapas do IBGE.

Esse processo foi por vezes repetidos até que se obtivesse uma união exata das bases com os dados corretos de CEP, bairro, distrito e os dados da SUSEP.

Com a análise dos dados o objetivo é entender os diferentes preços cobrados pelo mercado segurador na cidade de São Paulo e entender os fatores que contribuem para essa diferença de preços. Para facilitar a leitura e a organização dos dados foi utilizado o *software* R Studio versão 3.6, o *software* livre QGIS, e o Microsoft Excel para elaboração de tabelas e gráficos. No Capítulo 4 são apresentados as análises e resultados do trabalho.

4. Análises e Resultados

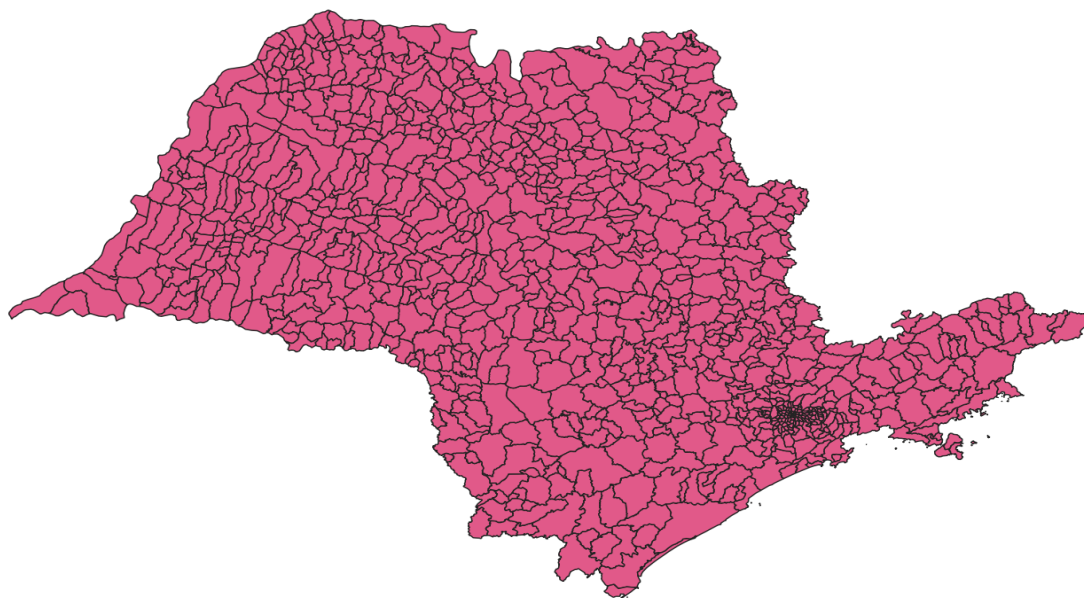
Nesse capítulo são analisados os dados da Autoseg. O arquivo utilizado foi o *arq_casco3_comp*, descrito no Quadro 1, pois é o único que contém o CEP. Nele tem-se informação sobre a exposição, número de sinistros de perda parcial e total, além de número de sinistros de roubo ou furto, através da chave: CEP, categoria tarifária, modelo, ano do veículo.

Os dados encontrados nas análises com as bases da Autoseg, estão coerentes com as estatísticas reais. É importante destacar que a base utilizada para a realização deste trabalho é pública, logo o entendimento das análises foi global e não particularizada, haja visto que as seguradoras utilizam apenas dados internos de seus segurados. Desta forma entende-se que se as seguradoras realizassem análises com as bases disponibilizadas pela SUSEP, mais a sua base interna, seria possível obter dados mais robustos e completos.

As bases foram carregadas no R semestre a semestre e para isso usou-se a função *read.csv*. Após esta etapa, as bases foram sumarizadas com o uso da função *aggregate*. Com isto as métricas foram calculadas: soma da exposição, soma de sinistros de roubo ou furto, soma dos sinistros de perda parcial ou total por CEP. Realizou-se o *merge* (união entre as bases) já sumarizada por CEP, com outra base que continha CEP e distrito de SP - esse procedimento foi realizado no Excel através da função *procv*. Após isso, a base foi sumarizada novamente, da mesma forma, só que desta vez, por distrito. As bases finais foram constituídas da soma de exposição, soma de sinistros de roubo ou furto e soma de sinistros perda parcial ou total por distrito da cidade de São Paulo em 2016, 2017 e 2018.

Entre as dificuldades encontradas no presente trabalho estão o fato da base da *autoseg* estar disponível por semestre e cada base, devido a chave utilizada, conter aproximadamente 10 milhões de registros totalizando 60 milhões. Outro problema foi que a base de dados da *autoseg* só tem CEP, não tem bairro e distrito, logo, houve-se a necessidade de uma base de dados de CEP, para fazer o *merge* e o *join*, e esse processo foi realizado manualmente com o uso do Excel. A base de CEP utilizada foi comprada em 2018 dos correios para a realização de um outro trabalho, o que foi útil, pois sem esses dados, seria necessário comprar uma. Um dado importante e que todo mês é criado um CEP novo no Brasil, por isso houve a necessidade de realizar uma conferência CEP a CEP, com cada bairro e distrito no Excel, afim de validar os dados e evitar possíveis erros.

A análise subsequente foi realizada no QGIS onde foram construídos e analisados os mapas. A base foi separada ano a ano para acompanhar a evolução da frequência de sinistros por distrito. O arquivo utilizado no QGIS (*shapefile*) foi baixado do site portal de mapas do IBGE e é composto por todos os distritos dos municípios do Estado de São Paulo (Mapa 1). A partir deste mapa foi filtrado o município de São Paulo que contém um total de 96 distritos e realizada a análise dos dados. Após a extração, foi realizada toda a análise subsequente no QGIS – mapas de frequência de roubo ou furto e frequência de perda parcial e perda total (que é definida pela soma do número de sinistro dividida pela soma da exposição) e a aplicação do *k-means* para clusterização dos distritos.



Mapa 1. Mapa dos distritos do Estado de São Paulo. Elaborado pela autora via *software* QGIS.

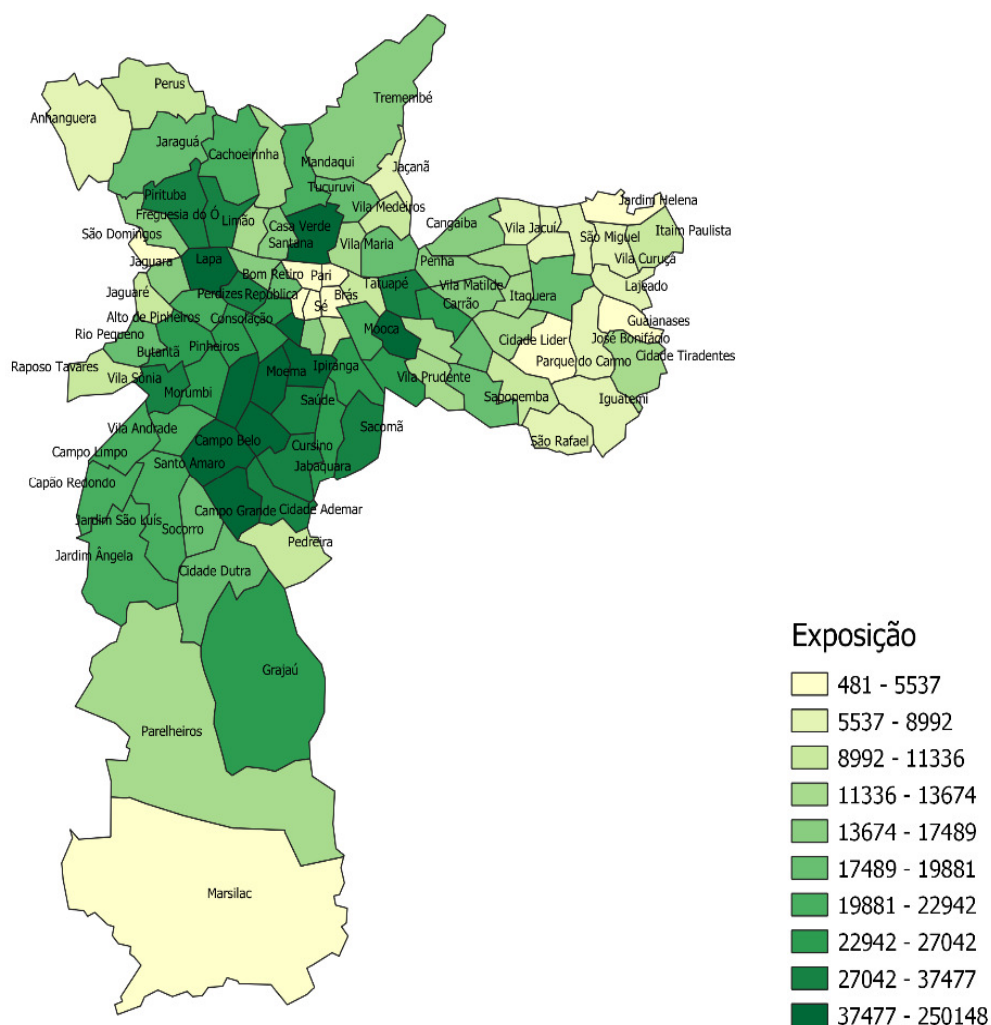
Na Tabela 1 tem-se frequências de roubo ou furto, perda total ou parcial em 2016, 2017 e 2018.

Tabela 2. Frequência de sinistros por ano

Ano	Roubo ou Furto	Perda parcial ou Total
2016	1,79%	4,60%
2017	1,55%	4,24%
2018	1,33%	3,50%

Tanto a frequência de roubo ou furto quanto a frequência de perda parcial e perda total diminuíram entre 2016 e 2018. A seguir o Mapa 2 com as exposições a risco de roubo ou furto por distrito em que se nota concentração na zona oeste, centro-sul e centro.

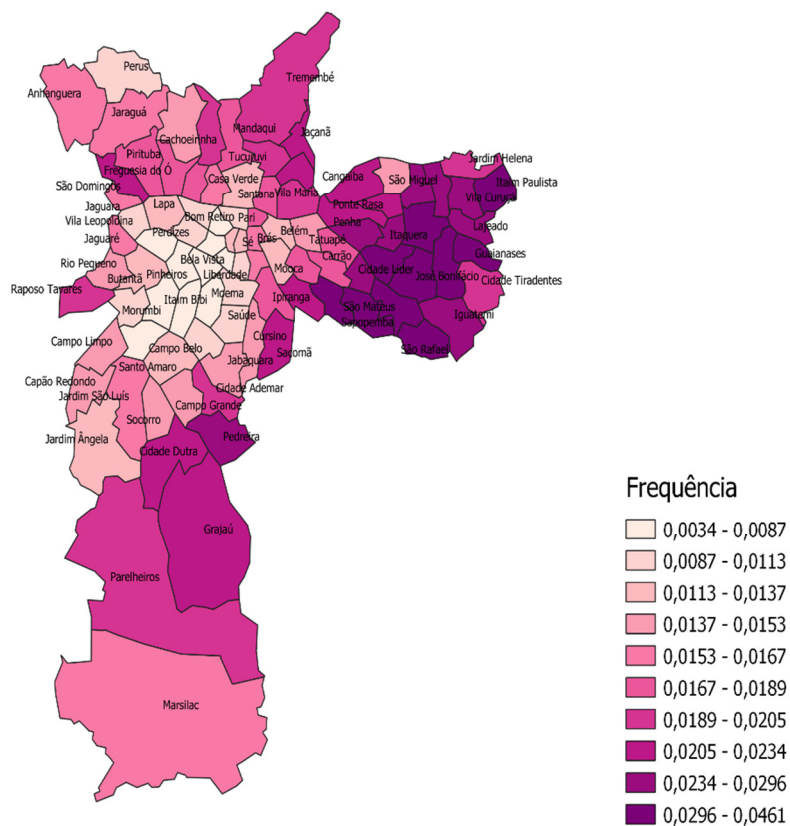
Exposição em São Paulo em 2016



Mapa 2. Exposição a risco de roubo ou furto em 2016. Elaborado pela autora via *software* QGIS.

Ao analisar o Mapa 3 que contém as frequências de roubo ou furto nos distritos do município de São Paulo em 2016, nota-se que os distritos das zonas leste e norte têm frequências elevadas de roubo ou furto enquanto as zona central e oeste apresentam menores índices.

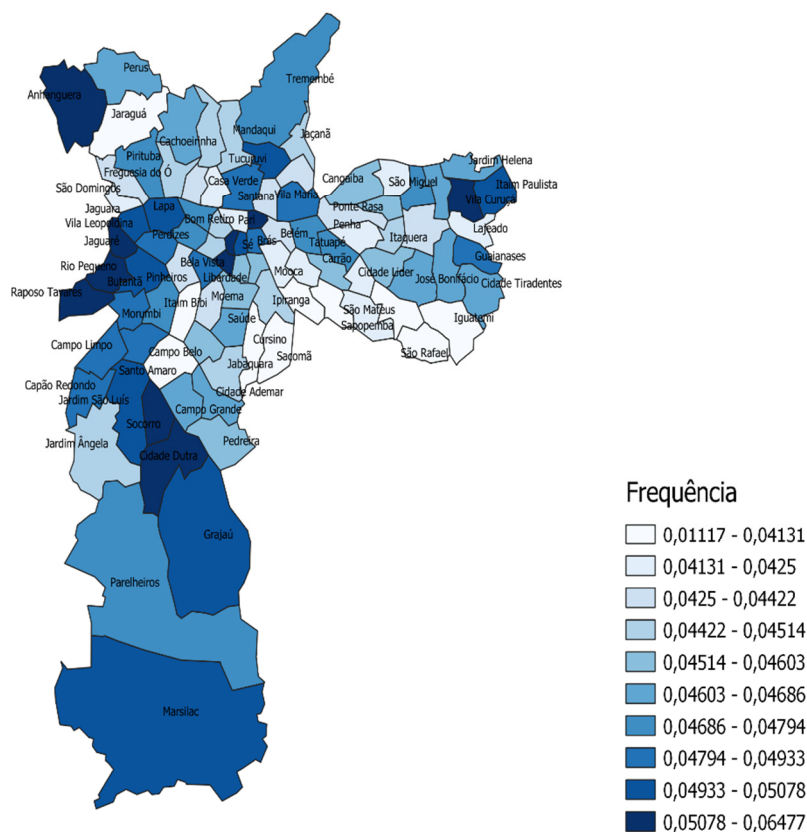
Frequência de roubo ou furto nos distritos do município de São Paulo em 2016



Mapa 3. Distritos dos Municípios de São Paulo a frequência de roubo ou furto em 2016.
Elaborado pela autora via *software* QGIS.

No Mapa 4 tem-se a frequência de perda parcial ou total em 2016. A zona leste não apresenta índices tão elevados e alguns distritos distribuídos em distintas zonas da capital apresentam frequências altas. Destaca-se a Lapa, Morumbi, Pinheiros, Butantã e Rio Pequeno.

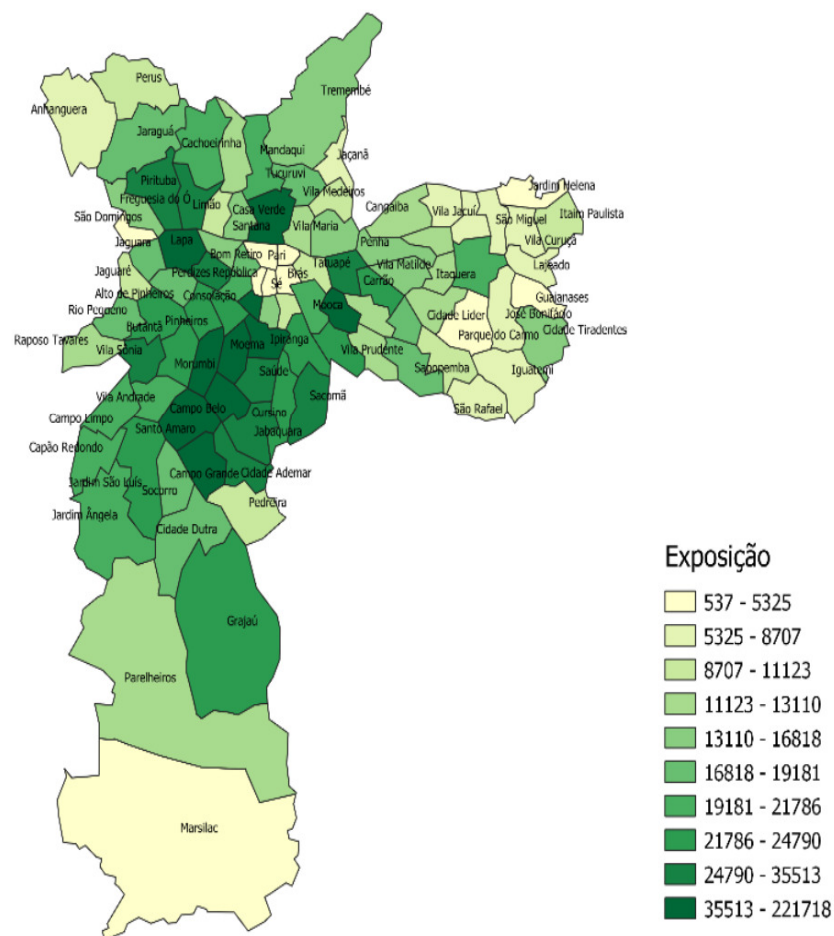
Frequência de PP ou PT nos distritos do município de São Paulo em 2016



Mapa 4. Distritos dos Municípios de São Paulo a frequência de PP ou PT em 2016.
Elaborado pela autora via *software* QGIS.

Os Mapas 5, 6 e 7 informam sobre a exposição e frequências em 2017. O primeiro mapa mostra as exposições por distrito em que se nota concentração na zona oeste, centro-sul e centro como ocorreu em 2016.

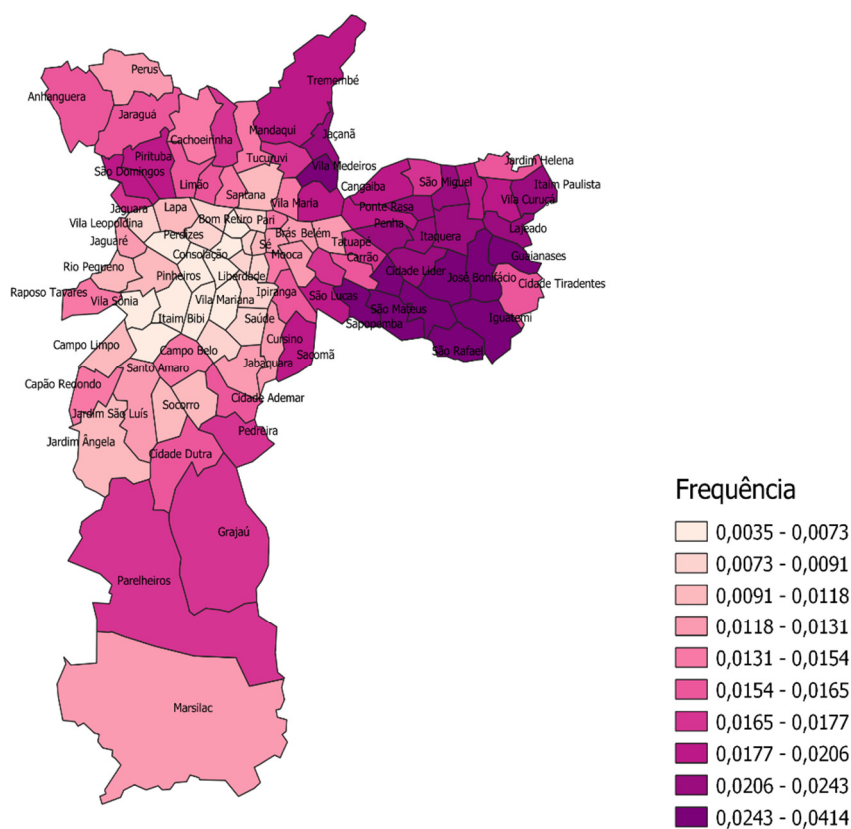
Exposição nos distritos do município de São Paulo em 2017



Mapa 5. Exposição a risco de roubo ou furto em 2017. Elaborado pela autora via *software* QGIS.

No Mapa 6 referente a frequência de roubo ou furto em 2017, nota-se como em 2016, que os distritos da zona leste (Sapopemba, Iguatemi, Parque do Carmo, Guaianases, São Mateus, Cidade Líder, Itaim Paulista), norte (Vila Medeiros, Jaçanã), têm frequências elevadas de roubo ou furto enquanto a zona central e oeste apresentam menores índices.

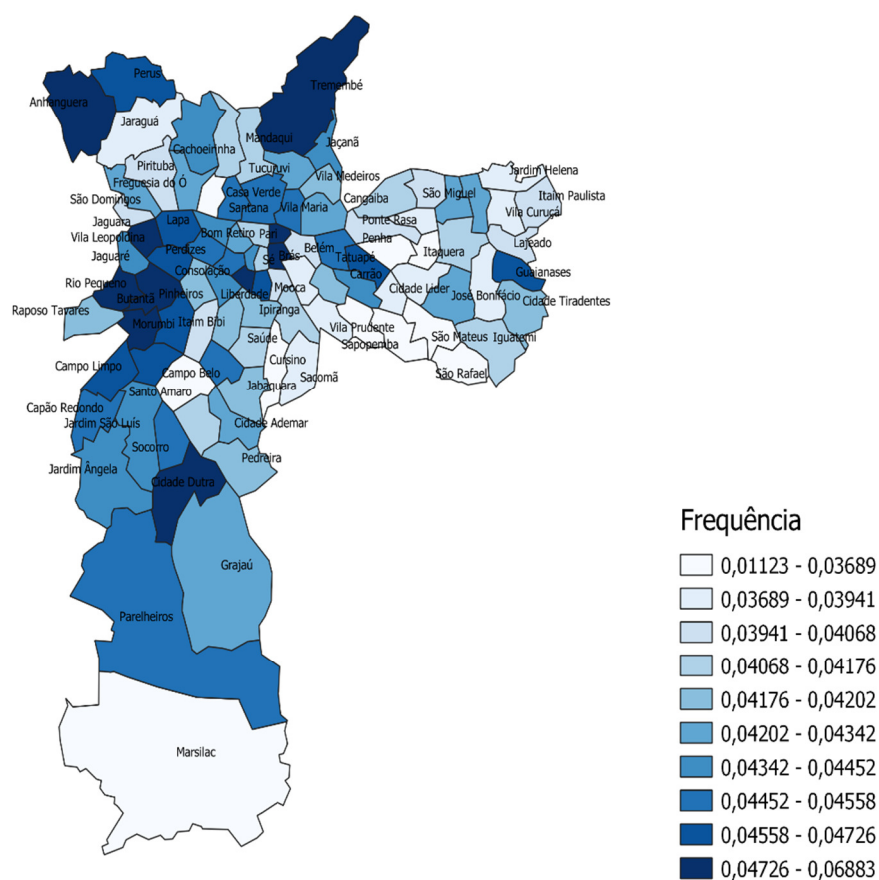
Frequência de roubo ou furto nos distritos do município de São Paulo em 2017



Mapa 6. Distritos dos Municípios de São Paulo com a frequência de roubo ou furto em 2017. Elaborado pela autora via *software* QGIS.

Já quanto a frequência de perda parcial ou total no ano de 2017 tem-se outro padrão conforme evidenciado no Mapa 7. A zona leste não apresenta índices tão elevados e alguns distritos distribuídos nas zonas da capital apresentam frequências altas ou menores. Destaca-se a zona oeste (Pinheiros, Butantã, Rio Pequeno).

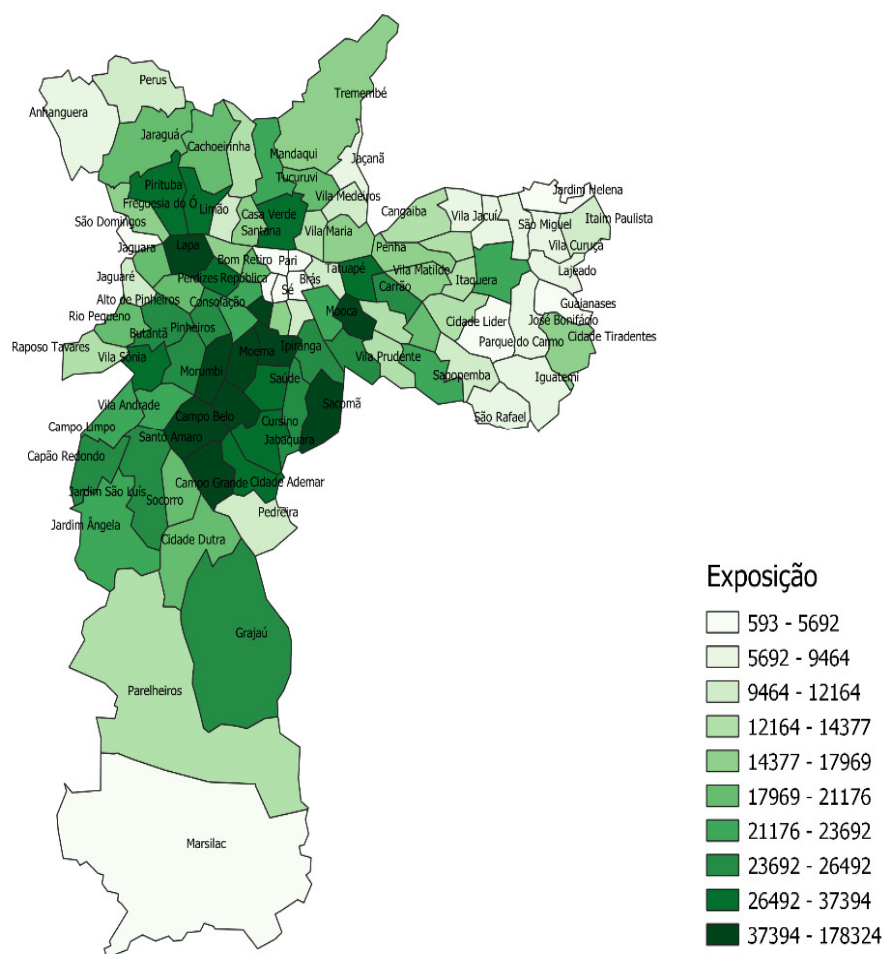
Frequência de PP ou PT nos distritos do município de São Paulo em 2017



Mapa 7. Distritos dos Municípios de São Paulo com a frequência de PP ou PT em 2017.
Elaborado pela autora via *software* QGIS.

A exposição em 2018 tem o mesmo padrão que 2016 e 2017 com concentração na zona central, centro-sul e oeste conforme Mapa 8.

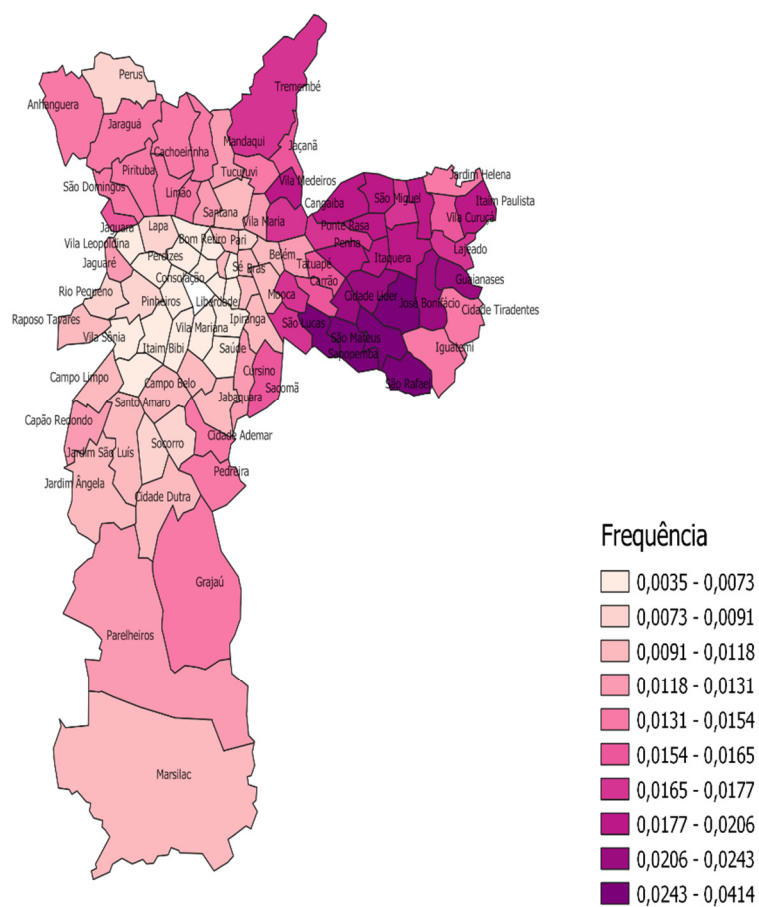
Exposição nos distritos do município de São Paulo em 2018



Mapa 8. Exposição a risco de roubo ou furto em 2018. Elaborado pela autora via *software* QGIS.

Além disso, em 2018, destaca-se o extremo sul da cidade de São Paulo que apresenta coloração menos intensa, porém a zona leste continua apresentando frequência elevada de roubo ou furto de acordo com o Mapa 9.

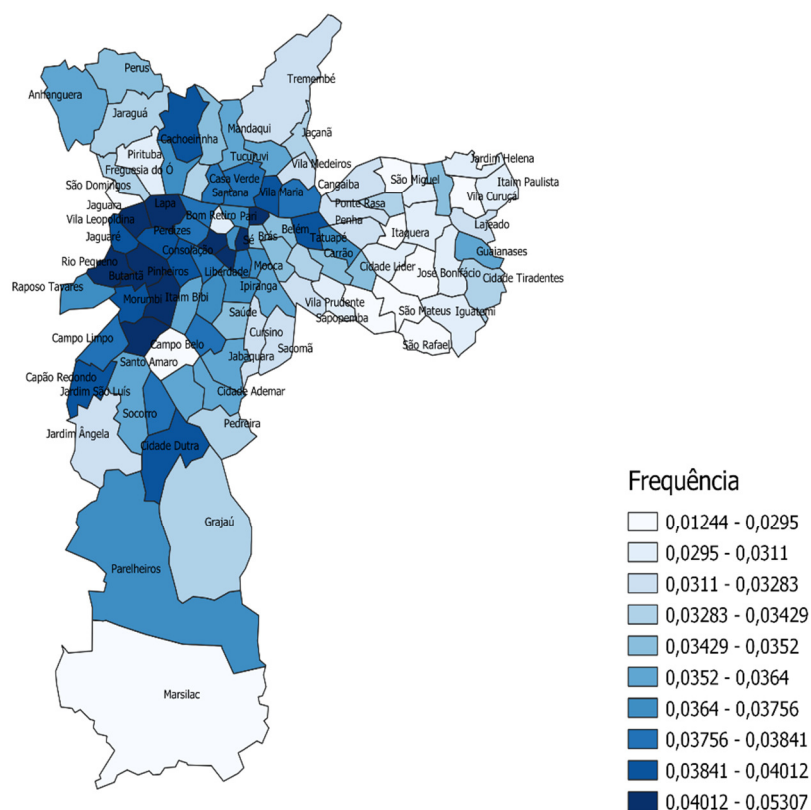
Frequência de roubo ou furto nos distritos do município de São Paulo em 2018



Mapa 9. Distritos dos Municípios de São Paulo com a frequência roubo ou furto em 2018.
Elaborado pela autora via *software* QGIS.

Quanto a frequência de perda parcial ou total tem-se outro padrão como ocorreu em 2017 e 2016 conforme visto no Mapa 10. A zona leste não apresenta índices tão elevados e alguns distritos distribuídos nas zonas da capital apresentam frequências altas ou menores. Destaca-se a Lapa, Morumbi, Pinheiros, Butantã e Rio Pequeno.

Frequência de PP ou PT nos distritos do município de São Paulo em 2018



Mapa 10. Distritos dos Municípios de São Paulo com a frequência de PP ou PT em 2018. Elaborado pela autora via *software* QGIS.

Após a análise dos mapas com as frequências de roubo ou furto e de frequência de perda total ou parcial foi aplicado o *k-means* para clusterização dos distritos. Como fora citado anteriormente a capital contém 96 distritos. Agrupar estes distritos segundo similaridade de frequências de roubo ou furto ou perda parcial e total pode ajudar a definir preços mais justos aos clientes e tornar as taxas mais viáveis para monitoramento pelas seguradoras.

Definiu-se 12 agrupamentos, esse agrupamento foi obtido através de testes, onde foram testados os números de *clusters* de 8 até 20, e 12 foi o número em que se obteve uma distribuição equilibrada entre os *clusters*, para que não se obtivesse uns com muitos e outros com poucos, da mesma forma calculou-se as frequências de roubo ou furto para os *clusters* definidos em 2016, 2017 e 2018. Os números dos *clusters* não significam que os mesmos distritos estão contidos neles ao longo do período analisado. Nota-se que os *clusters*

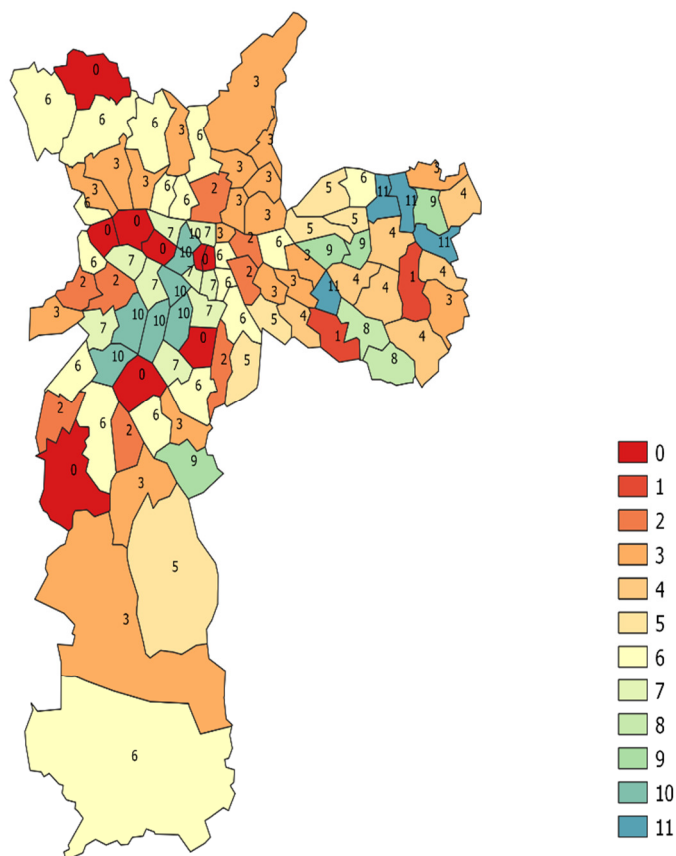
apresentam frequências distintas entre si. Os *clusters* para frequência de roubo ou furto foram elaborados pelo método *k-means* e estão apresentados na Tabela 3.

Tabela 3. Quantidade de distritos e frequências de roubo ou furto por *cluster* por ano

<i>Cluster</i>	<i>n</i>	<i>Freq. - 2016</i>	<i>Cluster</i>	<i>n</i>	<i>Freq. - 2017</i>	<i>Cluster</i>	<i>n</i>	<i>Freq. - 2018</i>
0	9	1,15%	0	6	2,87%	0	8	1,04%
1	2	3,43%	1	9	2,30%	1	7	0,92%
2	8	1,38%	2	10	1,78%	2	3	0,35%
3	20	1,96%	3	4	1,68%	3	7	1,34%
4	7	3,04%	4	2	3,94%	4	4	0,82%
5	6	2,23%	5	12	0,83%	5	7	0,49%
6	18	1,59%	6	5	1,63%	6	8	0,66%
7	9	0,89%	7	13	1,32%	7	11	1,95%
8	2	4,25%	8	7	1,56%	8	7	2,49%
9	4	2,50%	9	12	1,14%	9	12	1,66%
10	7	0,56%	10	7	2,02%	10	12	1,19%
11	4	2,69%	11	9	0,49%	11	10	1,48%
Total Geral	96	1,79%	Total Geral	96	1,55%	Total Geral	96	1,33%

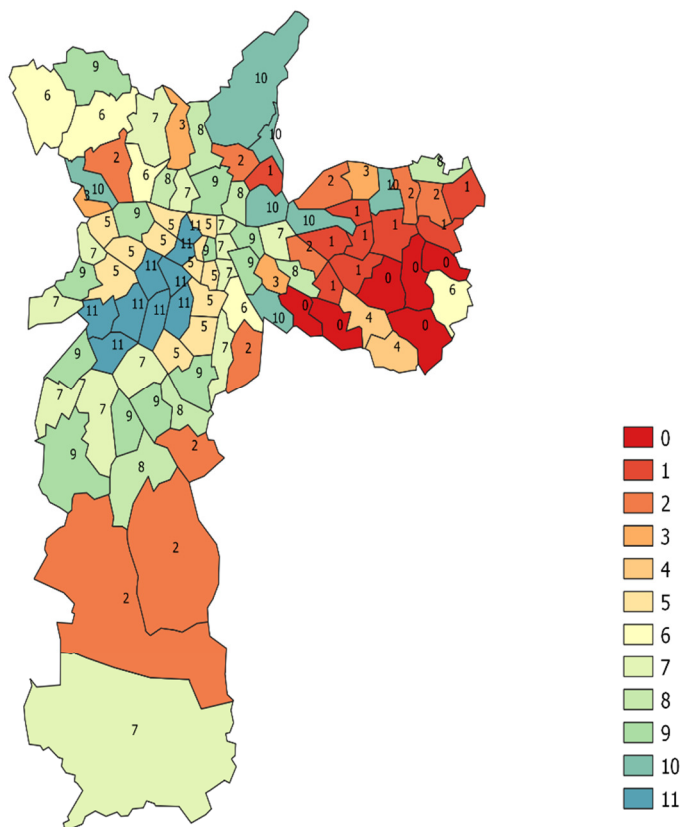
Nos Mapas de 11 a 16 os distritos que fazem parte do mesmo cluster têm a mesma cor e mesmo número. Uma vantagem do *k-means* além da redução de 96 distritos para 12 agrupamentos é a possibilidade de considerar a frequência de sinistro ao invés de utilizar somente as regiões ou sub-regiões. Se, por exemplo, um distrito tiver baixa frequência ele pode ser penalizado somente por estar próximo a outro de alta frequência. Após os mapas tem-se um quadro com os distritos que compõem os *cluster* e suas frequências de roubo ou furto.

Agrupamento de distritos do município de São Paulo via k-means para frequência de roubo ou furto (n = 12)



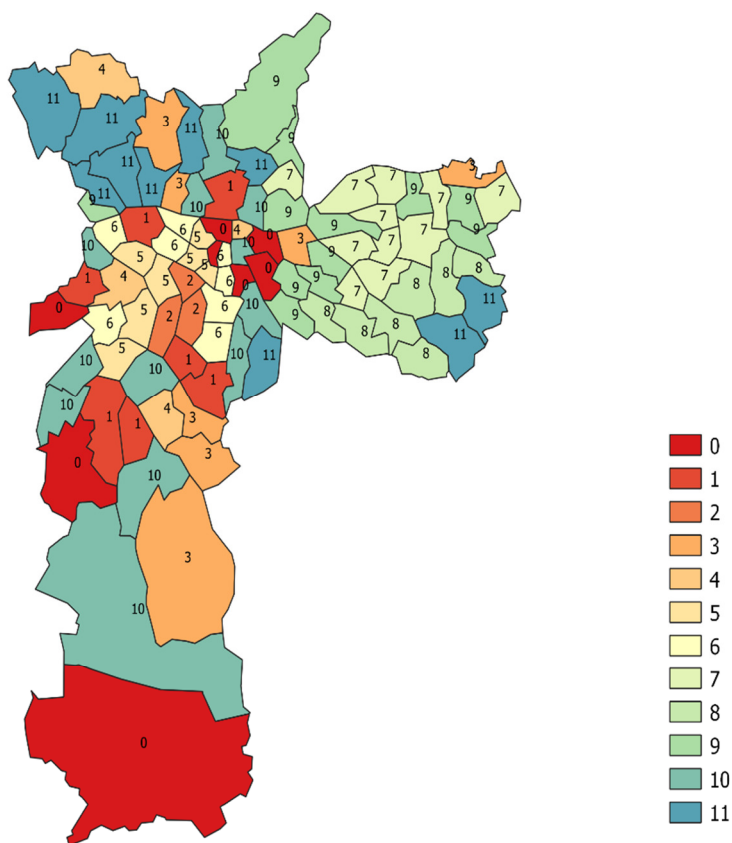
Mapa 11. Análise K-Means frequência de roubo ou furto em 2016. Elaborado pela autora via *software* QGIS.

Agrupamento de distritos do município de São Paulo via k-means para frequência de roubo ou furto (n = 12)



Mapa 12. Análise *k-means* frequência de roubo ou furto em 2017. Elaborado pela autora via *software* QGIS.

Agrupamento de distritos do município de São Paulo via k-means para frequência de roubo ou furto (n = 12)



Mapa 13. Análise *k-means* frequência de roubo ou furto em 2018. Elaborado pela autora via *software* QGIS.

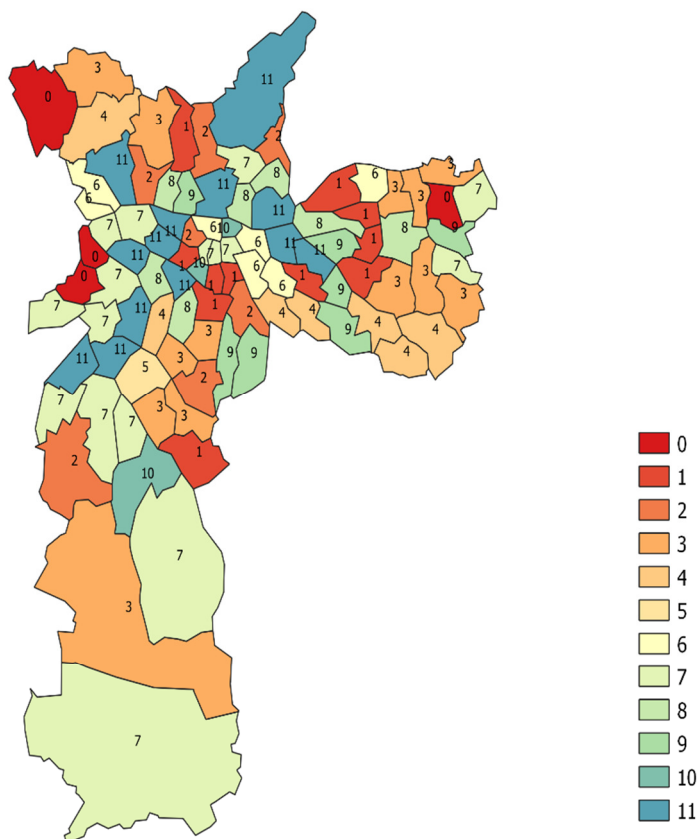
De acordo com a análise realizada na Tabela 3 e nos Mapas 11,12 e 13 para frequências de roubo ou furto, foi realizado um procedimento similar para frequência de perda parcial ou total. Na Tabela 4 a seguir tem-se as frequências calculadas.

Tabela 4. Quantidade de distritos e frequências de perda parcial ou total por *cluster* por ano

<i>Cluster</i>	<i>n</i>	<i>Freq. - 2016</i>	<i>Cluster</i>	<i>n</i>	<i>Freq. - 2017</i>	<i>Cluster</i>	<i>n</i>	<i>Freq. - 2018</i>
0	4	5,41%	0	2	2,74%	0	2	4,82%
1	11	4,53%	1	16	4,51%	1	5	2,71%
2	7	4,46%	2	2	6,44%	2	12	3,81%
3	13	4,65%	3	1	1,12%	3	11	3,97%
4	7	3,88%	4	6	3,80%	4	2	1,63%
5	1	1,12%	5	10	4,76%	5	1	5,31%
6	7	4,25%	6	12	3,99%	6	11	3,54%
7	16	5,01%	7	7	4,28%	7	11	3,68%
8	7	4,36%	8	19	4,17%	8	4	4,24%
9	7	4,16%	9	9	3,63%	9	12	2,99%
10	3	6,24%	10	3	5,32%	10	14	3,41%
11	13	4,79%	11	9	4,36%	11	11	3,21%
Total Geral	96	4,60%	Total Geral	96	4,24%	Total Geral	96	3,50%

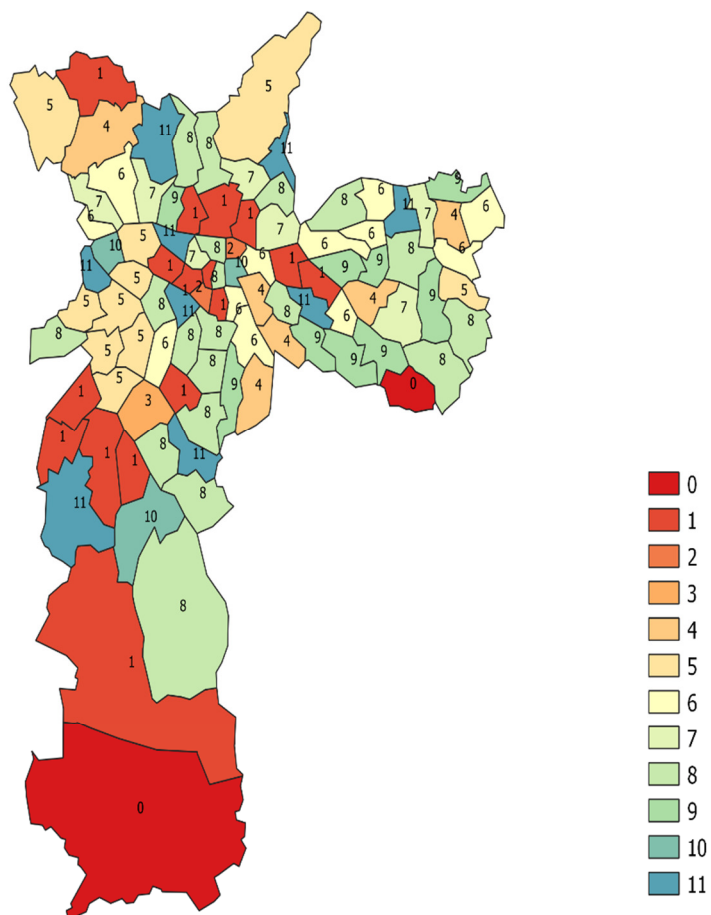
Nos Mapas de 14 a 16 pode-se visualizar os grupos em que distritos que fazem parte do mesmo *cluster* tem mesma cor e mesmo número.

Agrupamento de distritos do município de São Paulo via k-means para frequência PP ou PT (n = 12)



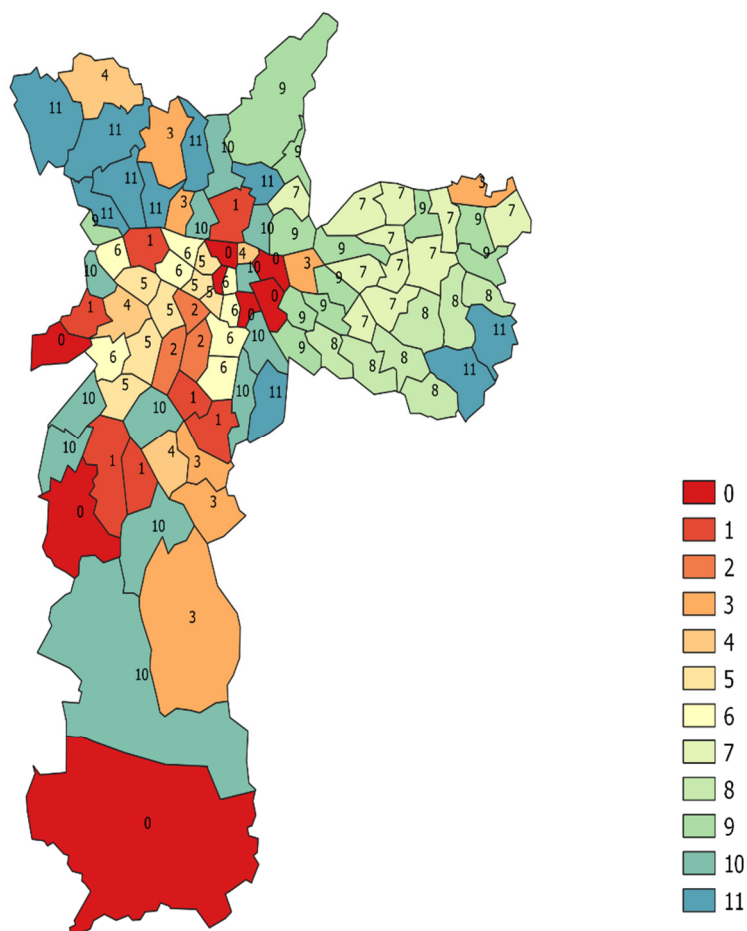
Mapa 14. Análise *k-means* frequência de perda parcial ou perda total em 2016. Elaborado pela autora via *software* QGIS.

Agrupamento de distritos do município de São Paulo via k-means para frequência PP ou PT (n = 12)



Mapa 15. K-Means frequência de perda parcial ou perda total 2017. Elaborado pela autora via *software* QGIS.

Agrupamento de distritos do município de São Paulo via k-means para frequência de roubo ou furto (n = 12)



Mapa 16. Análise k-means frequência de perda parcial ou perda total em 2018. Elaborado pela autora via *software* QGIS.

Nos Quadros de 1 a 6 encontram-se os distritos e seus *clusters* para roubo ou furto e para perda total ou parcial. Os *clusters* tem por finalidade demonstrar se as frequências de sinistros estão muito diferentes entre os clusters e próximas dentro dos *clusters*. Uma vez que isto ocorre, considera-se os clusters como bem definidos.

Foram definidos 12 *clusters* de distritos do município de São Paulo a partir da frequência de roubo ou furto e 12 *cluster* a partir da frequência de perda parcial ou perda total. O município contém 96 distritos e um agrupamento torna mais prático o monitoramento das taxas pela seguradora.

Em geral, distritos circunvizinhos apresentaram frequência de sinistros similares, o que torna razoável agrupá-los segundo a proximidade geográfica, e este comportamento se manteve no período analisado. No entanto, ao utilizar o *k-means* para criar os *clusters*, distritos de zonas distintas foram inseridos dentro de um mesmo grupo, o que evidencia que outros fatores influenciam no risco, como infraestrutura, entre outros.

Um aspecto interessante é o fato de que alguns distritos com taxas elevadas de roubo ou furto não apresentam altas frequências de perda parcial ou total. Por outro lado, alguns distritos com baixa frequência de roubo apresentam taxas elevadas de perda parcial ou total. Uma explicação para a maior incidência de roubo estaria na maior distância dos distritos ao centro. Distritos com esta característica geralmente tem menor policiamento ostensivo enquanto os distritos mais próximos ao centro têm maior aparato policial. Além disso, o trânsito no centro é mais intenso e de baixa velocidade média o que favorece pequenas colisões. Porém, somente um estudo mais detalhado para chegar a conclusões robustas.

Apesar de obter grupos com frequências de sinistros distintas entre si e similares dentro dos grupos, uma das desvantagens do *k-means* é que se deve definir inicialmente o número de *clusters*, além da sua performance em grandes bancos de dados não ser satisfatória.

A seguir os Quadros 3 a 8 com as frequências de sinistros dos *clusters* obtidos.

Quadro 3. Distritos e *clusters* para roubo ou furto em 2016

Cluster	Freq.	Distritos - Roubo ou Furto 2016						
0	1,15%	Jardim Ângela	Lapa	Perdizes	Perus	República	Santo Amaro	Saúde
		1,25%	1,18%	1,08%	1,13%	1,15%	1,20%	1,08%
		Sé	Vila Leopoldina					
		1,23%	1,06%					
1	3,43%	José Bonifácio	Sapopemba					
2	1,38%	3,29%	3,58%					
		Belém	Butantã	Capão Redondo	Cursino	Mooca	Rio Pequeno	Santana
		1,33%	1,33%	1,42%	1,47%	1,33%	1,40%	1,34%
		Socorro						
3	1,96%	1,46%						
		Água Rasa	Cachoeirinha	Carrão	Cidade Ademar	Cidade Dutra	Cidade Tiradentes	Freguesia do Ó
		1,78%	2,00%	1,85%	2,00%	2,08%	1,99%	1,85%
		Jaçanã	Jardim Helena	Parelheiros	Pari	Pirituba	Raposo Tavares	São Domingos
		2,06%	2,03%	2,03%	1,89%	1,88%	1,90%	2,06%
		Tremembé	Tucuruvi	Vila Formosa	Vila Guilherme	Vila Maria	Vila Medeiros	
4	3,04%	2,04%	2,04%	1,82%	1,89%	1,97%		
		Cidade Lder	Guaianases	Iguatemi	Itaim Paulista	Itaquera	Parque do Carmo	São Lucas
		3,09%	3,03%	2,93%	2,99%	3,17%	3,05%	3,06%
5	2,23%	Cangaíba	Grajaú	Penha	Ponte Rasa	Sacomã	Vila Prudente	
		2,21%	2,14%	2,25%	2,23%	2,34%	2,21%	
		Anhanguera	Brás	Brasilândia	Cambuci	Campo Grande	Campo Limpo	Casa Verde
6	1,59%	1,54%	1,59%	1,53%	1,54%	1,51%	1,51%	1,54%
		Ermelino Matarazzo	Ipiranga	Jabaquara	Jaguara	Jaguareé	Jaraguá	Jardim São Luís
		1,52%	1,74%	1,53%	1,60%	1,60%	1,62%	1,67%
		Limão	Mandaqui	Marsilac	Tatuapé			
		1,67%	1,72%	1,66%	1,53%			
7	0,89%	Alto de Pinheiros	Barra Funda	Bela Vista	Bom Retiro	Campo Belo	Liberdade	Pinheiros
		0,80%	0,87%	0,76%	0,89%	0,98%	0,99%	0,86%
		Vila Mariana	Vila Sônia					
		1,00%	0,87%					
8	4,25%	São Mateus	São Rafael					
		3,88%	4,61%					
9	2,50%	Artur Alvim	Pedreira	Vila Curuçá	Vila Matilde			
		2,57%	2,49%	2,52%	2,42%			
10	0,56%	Consolação	Itaim Bibi	Jardim Paulista	Moema	Morumbi	Santa Cecília	Vila Andrade
		0,57%	0,50%	0,34%	0,58%	0,61%	0,63%	0,69%
11	2,69%	Aricanduva	Lajeado	São Miguel	Vila Jacuí			
		2,73%	2,72%	2,69%	2,64%			

Quadro 4. Distritos e *clusters* para roubo ou furto em 2017

Cluster	Freq.	Distritos - Roubo ou Furto 2017						
		Guaianases	Iguatemi	José Bonifácio	Parque do Carmo	São Lucas	Sapopemba	
0	2,87%	2,72%	2,93%	2,87%	3,05%	2,69%	2,97%	
1	2,30%	Aricanduva	Artur Alvim	Cidade Líder	Itaim Paulista	Itaquera	Lajeado	Ponte Rasa
		2,50%	2,16%	2,38%	2,22%	2,34%	2,30%	2,16%
		Vila Matilde	Vila Medeiros					
2	1,78%	2,19%	2,48%					
		Cangaíba	Carrão	Grajaú	Parelheiros	Pedreira	Pirituba	Sacomã
		1,85%	1,77%	1,75%	1,75%	1,76%	1,78%	1,81%
3	1,68%	São Miguel	Tucuruvi	Vila Curuçá				
		1,78%	1,75%	1,85%				
		Água Rasa	Cachoeirinha	Ermelino Matarazzo	Jaguara			
4	3,94%	1,67%	1,70%	1,67%	1,69%			
		São Mateus	São Rafael					
		3,74%	4,14%					
5	0,83%	Alto de Pinheiros	Barra Funda	Bela Vista	Bom Retiro	Butantã	Campo Belo	Liberdade
		0,73%	0,79%	0,73%	0,90%	0,95%	0,84%	0,91%
		Perdizes	República	Saúde	Vila Leopoldina	Vila Mariana		
6	1,63%	0,88%	0,78%	0,88%	0,82%	0,76%		
		Anhanguera	Cidade Tiradentes	Freguesia do Ó	Ipiranga	Jaraguá		
		1,62%	1,61%	1,65%	1,63%	1,64%		
7	1,32%	Brás	Brasilândia	Cambuci	Capão Redondo	Casa Verde	Cursino	Jaguareé
		1,25%	1,34%	1,42%	1,37%	1,40%	1,26%	1,29%
		Jardim São Luís	Marsilac	Pari	Raposo Tavares	Santo Amaro	Tatuapé	
8	1,56%	1,24%	1,30%	1,32%	1,33%	1,32%	1,31%	
		Cidade Ademar	Cidade Dutra	Jardim Helena	Limão	Mandaqui	Vila Formosa	Vila Guilherme
		1,56%	1,59%	1,58%	1,58%	1,49%	1,59%	1,51%
9	1,14%	Belém	Campo Grande	Campo Limpo	Jabaquara	Jardim Ângela	Lapa	Mooca
		1,22%	1,17%	1,11%	1,20%	1,11%	1,02%	1,21%
		Perus	Rio Pequeno	Santana	Sé	Socorro		
10	2,02%	1,19%	1,16%	1,07%	1,10%	1,11%		
		Jaçanã	Penha	São Domingos	Tremembé	Vila Jacuí	Vila Maria	Vila Prudente
		2,09%	1,95%	1,95%	2,02%	2,06%	1,99%	2,06%
11	0,49%	Consolação	Itaim Bibi	Jardim Paulista	Moema	Morumbi	Pinheiros	Santa Cecília
		0,54%	0,38%	0,35%	0,45%	0,43%	0,59%	0,53%
		Vila Andrade	Vila Sônia					
		0,56%	0,60%					

Quadro 5. Distritos e *clusters* para roubo ou furto em 2018

Cluster	Freq.	Distritos - Roubo ou Furto 2018						
		Belém	Bom Retiro	Cambuci	Jardim Ângela	Marsilac	Mooca	Raposo Tavares
0	1,04 %	1,06 %	1,05 %	1,04 %	1,07 %	1,01 %	0,98 %	1,06 %
		República						
		1,01 %						
1	0,92 %	Campo Belo	Jabaquara	Jardim São Luís	Lapa	Rio Pequeno	Santana	Socorro
		0,96 %	0,92 %	0,95 %	0,88 %	0,90 %	0,93 %	0,87 %
2	0,35 %	Itaim Bibi	Jardim Paulista	Moema				
		0,37 %	0,31 %	0,37 %				
3	1,34 %	Brasilândia	Cidade Ademar	Grajaú	Jardim Helena	Limão	Pedreira	Tatuapé
		1,32 %	1,34 %	1,37 %	1,32 %	1,38 %	1,36 %	1,31 %
4	0,82 %	Butantã	Campo Grande	Pari	Perus			
		0,82 %	0,84 %	0,82 %	0,79 %			
5	0,49 %	Alto de Pinheiros	Bela Vista	Consolação	Morumbi	Pinheiros	Santa Cecília	Vila Andrade
		0,51 %	0,52 %	0,44 %	0,51 %	0,52 %	0,46 %	0,47 %
6	0,66 %	Barra Funda	Liberdade	Perdizes	Saúde	Sé	Vila Leopoldina	Vila Mariana
		0,65 %	0,70 %	0,65 %	0,66 %	0,66 %	0,68 %	0,66 %
		Vila Sônia						
7	1,95 %	0,60 %						
		Aricanduva	Artur Alvim	Cangaíba	Cidade Líder	Ermelino Matarazzo	Itaim Paulista	Itaquera
		2,17 %	1,86 %	1,82 %	2,13 %	1,82 %	2,00 %	1,96 %
8	2,49 %	Ponte Rasa	São Miguel	Vila Matilde	Vila Medeiros			
		2,02 %	1,85 %	1,84 %	1,94 %			
		Guaianases	José Bonifácio	Parque do Carmo	São Lucas	São Mateus	São Rafael	Sapopemba
9	1,66 %	2,26 %	2,32 %	2,57 %	2,59 %	2,63 %	2,55 %	2,49 %
		Água Rasa	Carrão	Jaçanã	Jaguara	Lajeado	Penha	Tremembé
		1,66 %	1,58 %	1,57 %	1,59 %	1,70 %	1,69 %	1,70 %
		Vila Curuçá	Vila Formosa	Vila Jacuí	Vila Maria	Vila Prudente		
10	1,19 %	1,64 %	1,62 %	1,77 %	1,74 %	1,67 %		
		Brás	Campo Limpo	Capão Redondo	Casa Verde	Cidade Dutra	Cursino	Ipiranga
		1,17 %	1,16 %	1,23 %	1,20 %	1,11 %	1,25 %	1,16 %
		Jaguareé	Mandaqui	Parelheiros	Santo Amaro	Vila Guilherme		
11	1,48 %	1,22 %	1,21 %	1,21 %	1,14 %	1,21 %		
		Anhanguera	Cachoeirinha	Cidade Tiradentes	Freguesia do Ó	Iguatemi	Jaraguá	Pirituba
		1,43 %	1,51 %	1,50 %	1,44 %	1,44 %	1,52 %	1,44 %
		Sacomã	São Domingos	Tucuruvi				
		1,54 %	1,48 %	1,52 %				

Quadro 6. Distritos e *clusters* perda parcial ou total em 2016

Cluster	Freq.	Distritos - Perda Parcial ou Total 2016						
0	5,41%	Anhanguera	Jaguaré	Rio Pequeno	Vila Curuçá			
		5,30%	5,36%	5,47%	5,51%			
1	4,53%	Artur Alvim	Cachoeirinha	Cambuci	Cangaíba	Cidade Líder	Consolação	
		4,49%	4,50%	4,53%	4,52%	4,57%	4,51%	
		Liberdade	Pedreira	Ponte Rasa	Vila Formosa	Vila Mariana		
		4,53%	4,52%	4,55%	4,55%	4,53%		
2	4,46%	Freguesia do Ó	Ipiranga	Jabaquara	Jaçanã	Jardim Ângela	Mandaqui	Santa Cecília
		4,47%	4,45%	4,45%	4,46%	4,48%	4,44%	4,46%
3	4,65%	Brasilândia	Campo Belo	Campo Grande	Cidade Ademair	Cidade Tiradentes	Jardim Helena	José Bonifácio
		4,69%	4,60%	4,65%	4,62%	4,61%	4,67%	4,69%
		Parelheiros	Parque do Carmo	Perus	São Miguel	Saúde	Vila Jacuí	
		4,69%	4,63%	4,61%	4,66%	4,61%	4,69%	
4	3,88%	Iguatemi	Itaim Bibi	Jaraguá	São Lucas	São Mateus	São Rafael	Vila Prudente
		3,95%	3,84%	3,98%	3,82%	3,75%	3,91%	3,94%
5	1,12%	Santo Amaro						
		1,12%						
6	4,25%	Água Rasa	Belém	Bom Retiro	Ermelino Matarazzo	Jaguara	Mooca	São Domingos
		4,22%	4,28%	4,25%	4,25%	4,23%	4,25%	4,29%
7	5,01%	Brás	Butantã	Capão Redondo	Grajaú	Guaianases	Itaim Paulista	Jardim São Luís
		4,91%	4,99%	4,90%	5,03%	4,91%	4,99%	5,08%
		Lapa	Marsilac	Raposo Tavares	República	Sé	Socorro	Tucuruvi
		5,06%	4,99%	5,08%	5,12%	5,03%	5,10%	5,00%
		Vila Leopoldina	Vila Sônia					
		4,98%	4,93%					
8	4,36%	Itaquera	Limão	Moema	Penha	Pinheiros	Vila Guilherme	Vila Medeiros
		4,34%	4,35%	4,41%	4,40%	4,34%	4,39%	4,33%
9	4,16%	Aricanduva	Casa Verde	Cursino	Lajeado	Sacomã	Sapopemba	Vila Matilde
		4,18%	4,20%	4,12%	4,18%	4,11%	4,20%	4,14%
10	6,24%	Bela Vista	Cidade Dutra	Pari				
		6,40%	6,48%	5,85%				
11	4,79%	Alto de Pinheiros	Barra Funda	Campo Limpo	Carrão	Jardim Paulista	Morumbi	Perdizes
		4,82%	4,74%	4,82%	4,77%	4,82%	4,76%	4,77%
		Pirituba	Santana	Tatuapé	Tremembé	Vila Andrade	Vila Maria	
		4,74%	4,88%	4,72%	4,76%	4,84%	4,85%	

Quadro 7. Distritos e *clusters* perda parcial ou total em 2017

Cluster	Freq.	Distritos - Perda Parcial ou Total 2017						
0	2,74%	Marsilac	São Rafael					
		2,61%	2,87%					
1	4,51%	Campo Belo	Campo Limpo	Capão Redondo	Carrão	Casa Verde	Consolação	Jardim São Luís
		4,53%	4,57%	4,51%	4,58%	4,56%	4,48%	4,44%
		Liberdade	Parelheiros	Perdizes	Perus	República	Santana	Socorro
		4,56%	4,49%	4,47%	4,61%	4,45%	4,46%	4,47%
		Tatuapé	Vila Guilherme					
		4,52%	4,45%					
2	6,44%	Bela Vista	Pari					
		6,88%	6,00%					
3	1,12%	Santo Amaro						
		1,12%						
4	3,80%	Cidade Líder	Jaraguá	Mooca	Sacomã	Vila Curuçá	Vila Prudente	
		3,82%	3,82%	3,79%	3,82%	3,82%	3,75%	
5	4,76%	Alto de Pinheiros	Anhanguera	Butantã	Guaianases	Lapa	Morumbi	Rio Pequeno
		4,67%	4,89%	4,86%	4,65%	4,64%	4,72%	4,97%
		Tremembé	Vila Andrade	Vila Sônia				
		4,74%	4,70%	4,80%				
6	3,99%	Aricanduva	Belém	Cambuci	Ermelino Matarazzo	Ipiranga	Itaim Bibi	Itaim Paulista
		3,94%	3,98%	3,99%	4,07%	4,07%	3,95%	3,97%
		Itaim Paulista	Jaguara	Lajeado	Penha	Pirituba	Ponte Rasa	
		3,97%	3,99%	4,02%	4,01%	4,00%	3,93%	
7	4,28%	Freguesia do Ó	Parque do Carmo	Santa Cecília	São Domingos	São Miguel	Tucuruvi	Vila Maria
		4,24%	4,28%	4,28%	4,26%	4,31%	4,31%	4,30%
8	4,17%	Água Rasa	Bom Retiro	Cachoeirinha	Campo Grande	Cangaíba	Cidade Tiradentes	Grajaú
		4,19%	4,18%	4,17%	4,17%	4,12%	4,19%	4,21%
		Iguatemi	Itaquera	Jabaquara	Mandaqui	Moema	Pedreira	Pinheiros
		4,13%	4,15%	4,19%	4,14%	4,19%	4,18%	4,19%
		Raposo Tavares	Saúde	Sé	Vila Mariana	Vila Medeiros		
		4,19%	4,15%	4,13%	4,20%	4,19%		
9	3,63%	Artur Alvim	Cursino	Jardim Helena	José Bonifácio	Limão	São Lucas	São Mateus
		3,67%	3,66%	3,71%	3,71%	3,67%	3,64%	3,43%
		Sapopemba	Vila Matilde					
		3,63%	3,60%					
10	5,32%	Brás	Cidade Dutra	Vila Leopoldina				
		5,34%	5,58%	5,05%				
11	4,36%	Barra Funda	Brasilândia	Cidade Ademar	Jaçanã	Jaguará	Jardim Ângela	Jardim Paulista
		4,38%	4,42%	4,32%	4,35%	4,39%	4,35%	4,36%
		Vila Formosa	Vila Jacuí					
		4,35%	4,34%					

Quadro 8. Distritos e *clusters* perda parcial ou total em 2018

Cluster	Freq.	Distritos - Perda Parcial ou Total 2018						
0	4,82%	Morumbi	Vila Leopoldina					
		5,06%	4,58%					
1	2,71%	Artur Alvim	Jaguara	Parque do Carmo	São Mateus	São Rafael		
		2,80%	2,63%	2,71%	2,76%	2,64%		
2	3,81%	Barra Funda	Campo Belo	Campo Limpo	Capão Redondo	Casa Verde	Jardim Paulista	Liberdade
		3,84%	3,81%	3,83%	3,86%	3,76%	3,78%	3,83%
		Perdizes	Raposo Tavares	Santana	Socorro	Vila Maria		
		3,84%	3,75%	3,83%	3,78%	3,76%		
3	3,97%	Alto de Pinheiros	Brasilândia	Butantã	Cidade Dutra	Jaguareé	Pari	Pinheiros
		3,96%	3,90%	4,04%	3,92%	3,94%	4,02%	3,90%
		Sé	Tatuapé	Vila Guilherme	Vila Sônia			
		4,07%	3,99%	3,92%	4,00%			
4	1,63%	Marsilac	Santo Amaro					
		2,02%	1,24%					
5	5,31%	Bela Vista						
		5,31%						
6	3,54%	Anhanguera	Cachoeirinha	Cidade Ademar	Ipiranga	Itaim Bibi	Jabaquara	Jardim São Luís
		3,55%	3,52%	3,55%	3,57%	3,57%	3,59%	3,52%
		Mandaqui	Mooca	Perus	São Miguel			
7	3,68%	Bom Retiro	Cambuci	Campo Grande	Carrão	Freguesia do Ó	Guaianases	Moema
		3,65%	3,72%	3,63%	3,69%	3,66%	3,64%	3,73%
		Parelheiros	República	Tucuruvi	Vila Mariana			
8	4,24%	Consolação	Lapa	Rio Pequeno	Vila Andrade			
		4,34%	4,21%	4,19%	4,21%			
9	2,99%	Cidade Líder	Ermelino Matarazzo	Iguatemi	Itaim Paulista	Itaquera	Jardim Helena	José Bonifácio
		3,01%	2,88%	3,00%	2,97%	2,96%	3,08%	3,05%
		Santa Cecília	São Lucas	Sapopemba	Vila Curuçá	Vila Jacuí		
10	3,41%	Água Rasa	Aricanduva	Belém	Brás	Cidade Tiradentes	Grajaú	Jaçanã
		3,33%	3,47%	3,44%	3,47%	3,41%	3,43%	3,40%
		Jaraguá	Limão	Pedreira	Ponte Rasa	São Domingos	Saúde	Vila Formosa
11	3,21%	Cangaíba	Cursino	Jardim Ângela	Lajeado	Penha	Pirituba	Sacomã
		3,18%	3,19%	3,19%	3,19%	3,28%	3,11%	3,25%
		Tremembé	Vila Matilde	Vila Medeiros	Vila Prudente			
		3,28%	3,24%	3,13%	3,24%			

No Quadro 3 tem-se os distritos e *clusters* para roubo ou furto em 2016, no qual pode-se observar a padronização dada pela escolha dos 12 *clusters*, no qual existe um equilíbrio no número de distritos dentro dos *clusters*, bem como dentro dos *clusters* frequências próximas (*Cluster 2*, frequência 1,38% e entre os bairros varia de 1,33% a 1,46%), e entre os *clusters* frequências distintas (*Cluster 0*, frequência 1,15%; *Cluster 1*, frequência 3,43%). Temos para o Quadro 4 a 8 a mesma interpretação do Quadro 3.

Em geral, é razoável considerar os distritos circunvizinhos com frequência de sinistros similares na construção de regiões de risco. No entanto, ao utilizar o algoritmo o *k-means* para criar os *clusters*, nota-se que distritos de zonas distintas foram inseridos dentro de um mesmo grupo, o que evidencia que outros aspectos influenciam no risco. Não necessariamente um distrito que está próximo a outro de alta frequência de sinistro apresenta o mesmo comportamento não se justificando, portanto, um preço elevado cobrado pelas seguradoras para tais localidades. Ao configurar região de risco tendo como critério principal

a proximidade geográfica acaba-se por penalizar um distrito de baixa risco pelo de fato de estarem próximos geograficamente a outro de alto risco. Os mapas e quadros evidenciam isto e possibilitam iniciar um estudo mais detalhado sobre o tema.

Após a apresentação das análises e resultados encontrados com o estudo, no Capítulo 5 tem-se algumas considerações finais.

5. Considerações Finais

O número de roubos e furtos de veículos impactam diretamente no preço dos seguros. Este número, por sua vez, varia regionalmente, sendo, geralmente maior na capital e região metropolitana do que no interior do Estado. Dessa forma, para que uma seguradora avalie o risco adequadamente é necessário ter informação sobre a frequência de roubos ou furtos. Além disso a frequência de perda total e perda parcial podem ser estudadas, uma vez que também impactam no preço final do seguro ao consumidor.

O valor do seguro considera o perfil do segurado, as coberturas contratadas e os sinistros computados pelas seguradoras. Quanto maior o risco, maior será o prêmio. Caso o segurado esteja numa região ou distrito com alto índice de roubo de veículo maior probabilidade de sofrer um roubo ou furto, maior o risco e maior será o prêmio.

A diferença entre as frequências de roubo ou furto e de perda parcial ou perda total explicam porque muitas vezes os preços praticados pelas seguradoras variam tanto dependendo do CEP informado, inclusive variando entre as próprias seguradoras. Ao utilizar as informações da sua própria base de clientes e da SUSEP ela avalia o risco de sinistro de roubo ou perda parcial, por exemplo, e se o cliente reside numa região com maior risco maior será o prêmio cobrado. Nota-se também que distritos ao longo de 2016, 2017 e 2018 apresentam taxas mais elevadas de roubo mantiveram o padrão assim como perda parcial.

Um aspecto interessante é o fato de que alguns distritos com taxas elevadas de roubo ou furto não apresentam altas frequências de perda parcial ou total. Por outro lado, alguns distritos com baixa frequência de roubo apresentam taxas elevadas de perda parcial ou total. Uma explicação para a maior incidência de roubo estaria na maior distância dos distritos ao centro. Distritos com esta característica geralmente tem menor policiamento ostensivo enquanto os distritos mais próximos ao centro têm maior aparato policial. Além disso, o trânsito no centro é intenso e de baixa velocidade média o que favorece pequenas colisões. De qualquer sorte, é necessário investigar os dados para chegar a conclusões se isto é um padrão ou só o acaso.

Dessa forma, no presente trabalho foi apresentado 12 *clusters* de distritos do município de São Paulo obtidos a partir do algoritmo *k-means* considerando a frequência de roubo ou furto e 12 cluster considerando a frequência de perda parcial ou perda total. O

município contém 96 distritos e um agrupamento torna viável o monitoramento das taxas pela seguradora.

Em geral, distritos circunvizinhos apresentaram frequência de sinistro similares, o que torna razoável agrupá-los segundo a proximidade geográfica, e este comportamento se manteve no período analisado. No entanto, ao utilizar o *k-means* para criar os *clusters*, distritos de zonas distintas foram inseridos dentro de um mesmo grupo, o que evidencia que outros aspectos influenciam no risco.

Para trabalhos futuros é possível utilizar outros algoritmos de agrupamento, ou aplicar técnicas que ajudem a definir o número de *cluster* e então aplicar o *k-means*. Outra recomendação é uma extensão do estudo para outras capitais, regiões metropolitanas ou até o país, uma vez que há regiões que apresentam similaridade quanto as frequências de sinistro de roubo ou furto e perda parcial ou total, ou seja, criar clusters considerando frequências de roubo ou furto e perda parcial ou total simultaneamente, além de considerar variáveis relacionadas a infraestrutura dos distritos, velocidade média das vias, renda, etc. Outra possibilidade é testar os *clusters* aqui obtidos e checar se estes são úteis na precificação.

6. Referências Bibliográficas

Análise Multivariada: FIPECAFI- Fundação Instituto de Pesquisa Contábeis, Atuariais e Financeiras; Luis J.Corrar, Edilson Paulo, José Maria Dias Filho (coordenadores_ - São Paulo: Atlas, 2007.

BARANAUSKAS, Jose Augusto. Clustering Agrupamento. < Disponível em: <http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Clustering.pdf>> 20 de Ago de 2020.

BOGORNY, Vânia. Formação de agrupamentos: conceitos básicos e algoritmos. < Disponível em: <http://www.inf.ufsc.br/~vania/teaching/INE5644/clustering1.ppt>> 22 de Out de 2020.

Em abril, furto de veículos caiu 49% e roubo 45%, no Estado de São Paulo <Disponível em: > <https://www.revistacobertura.com.br/2020/06/26/em-abril-furto-de-veiculos-caiu-49-e-roubo-45-no-estado-de-sao-paulo/>> acesso em: 10 de Ago de 2020.

FUNENSEG. <Disponível em: <https://www.ens.edu.br/>> acesso em: 10 de Ago de 2020.

Secretária de segurança pública do Estado de São Paulo. <Disponível em: <http://www.ssp.sp.gov.br/>> acesso em: 10 de Ago de 2020.

Guia de orientação e defesa do segurado.<Disponível em: https://www2.susep.gov.br/download/cartilha/cartilha_susep2e.pdf> acesso em: 10 de Ago de 2020.

Introdução Básica à Clusterização. HONDA, Hugo (2017).<Disponível em: https://lamfo-unb.github.io/2017/10/05/Introducao_basica_a_clusterizacao/> acesso em: 19 de Ago de 2020.

R7. SP tem 1 milhão de carros roubados e furtados no Estado em cinco anos. <. Disponível em: <https://noticias.r7.com/sao-paulo/sp-tem-1-milhao-de-carros-roubados-e-furtados-no-estado-em-cinco-anos-08032020>> acesso em: 10 de ago de 2020.

Práticas de geoprocessamento em QGIS / Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável. 2. ed. - Belo Horizonte: Semad, 2019.

Ministério da Infraestrutura do governo federal <. Disponível em:

<https://www.gov.br/acessoainformacao/pt-br/assuntos/relatorios-dados>> acesso em: 10 de ago de 2020.

FERREIRA, Paulo Pereira. Modelos de precificação e ruína para seguros de curto prazo – Rio de Janeiro: FUNENSEG, 2002.

MINUTOSEGUROS. O que influencia no valor do seguro auto? < Disponível em:

<https://www.minutoseguros.com.br/blog/o-que-influencia-no-valor-do-seguro-auto/#:~:text=%C3%8Dndice%20de%20roubo%20dos%20ve%C3%ADculos,cidade%2C%20pois%20pode%20haver%20altern%C3%A2ncias.>> > acesso em: 20 de ago de 2020

NUNES, Diogo Henriques Freitas. Perda Parcial. <.Disponível em:

<https://www.minutoseguros.com.br/perguntas-frequentes/seguro-auto/o-que-e-perda-parcial#:~:text=Perda%20parcial%20%C3%A9%20quando%20o,valor%20indicado%20na%20Tabela%20Fipe.&text=Neste%20caso%2C%20a%20seguradora%20considera,devolv e%20o%20pre%C3%A7o%20Fipe%20integral.>> > acesso em: 10 de ago de 2020.

Um breve estudo sobre o algoritmo K-means. Dissertação – Coimbra , 2016.

SANTANA, Felipe. Entenda o Algoritmo K-means e Saiba como Aplicar essa Técnica.

<.Disponível em:<https://minerandodados.com.br/entenda-o-algoritmo-k-means/>> > acesso em: 21 de Out. de 2020.

Sinistro: O que é e o que fazer nessa situação<.Disponível em:<

<https://www.meuportoseguro.com.br/meu-carro/seguro-auto/sinistro-o-que-e-e-o-que-fazer-nessa-situacao/>> acesso em: 10 de ago de 2020.

Seguro de carro. <.Disponível em:<https://quatorrodas.abril.com.br/auto-servico/seguro-do-carro-esta-carro-veja-11-dicas-para-deixa-lo-mais-em-conta/>> acesso em: 20 de ago de 2020.

Dados Estatísticos do Estado de São Paulo.< Disponível em: <https://www.ssp.sp.gov.br/estatistica/mapas.aspx>> acesso em: 20 de ago de 2020.

TUDOSOBRESEGUROS.Fundamentos do Seguro <.Disponível em:

<https://www.tudosobreseguros.org.br/fundamentos-do-seguro/#:~:text=Assim%2C%20estritamente%20falando%2C%20risco%20%C3%A9,data%20incerta%2C%20como%20a%20morte.>> > acesso em: 10 de ago de 2020.

Os catorze bairros onde mais se furtam carros em São Paulo. < Disponível em: <https://vejasp.abril.com.br/cidades/bairros-mais-furtos-carros-sp/>> acesso em: 20 de ago de 2020.

OLIVEIRA, Roberto. Explicando o Algoritmo Iterativo K-Means Clustering<.Disponível em: <https://abracd.org/explicando-o-algoritmo-iterativo-k-means-clustering/>> acesso em: 25 de ago de 2020.

SANTOS, Alex da Silva. Introdução ao ambiente SIG QGUS <.Disponível em: http://geoftp.ibge.gov.br/metodos_e_outros_documentos_de_referencia/outros_documentos_tecnicos/introducao_sig_qgis/Introducao_ao_ambiente_SIG_QGIS_2edicao.pdf acesso em: 25 de ago de 2020.

VEJASP.Bairros com mais furtos de carros em SP. <Disponível em: <https://vejasp.abril.com.br/cidades/bairros-mais-furtos-carros-sp/>> acesso em: 10 de Ago de 2020.

7. Anexos

Abaixo está descrito o código no R utilizado para sumarizar as bases da Autoseg. Lembrando que antes a após esse processo, utilizou-se o Excel para tratamento e ordenação dos dados, e ao final utilizou-se o QGIS para fazer os mapas e as análises do k-means.

```
#=====
```

```
#TCC Bárbara - Análise ....
```

```
##fonte:Autoseg,
```

```
##=====
```

```
#C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC  
II/Trabalho/bases/Autoseg2014A/AjudaAutoseg.html
```

```
###Objetivo
```

```
#O Sistema AUTOSEG permite realizar consultas referentes a dados estatísticos do Seguro  
#de Automóveis. Tais consultas podem ser realizadas on-line, ou através de base de dados  
#no formato Access, disponível para download no mesmo site.
```

```
#As informações apresentadas
```

```
#foram obtidas a partir dos arquivos enviados semestralmente pelas companhias  
#seguradoras, os quais incluem dados referentes a apólices vigentes e sinistros ocorridos.
```

```
#O Sistema fornece informações sobre número de veículos expostos, prêmio médio,  
#importância segurada média, número de sinistros e valor de indenizações, classificadas de  
#acordo com categoria, modelo e ano do veículo, região, cidade ou CEP de circulação, e  
#perfil do segurado.
```

```
#Os códigos de modelos, são os da codificação padronizada da tabela FIPE.
```

```
###Consultas On-line
```

```
#Existem atualmente duas consultas disponibilizadas on-line:
```

```
#Categoria Tarifária/Região/Modelo/Ano/Sexo/Idade; e
```

```
#Categoria Tarifária/Cidade/Modelo/Ano.
```

```
#A característica principal da consulta on-line é que a mesma pode ser feita por períodos,  
com dados disponíveis desde 01/07/2006.
```

```
###Base para Download
```

```
#Além da consulta on-line, é disponibilizada para download uma base de dados no formato  
#Access, com os dados do último envio semestral realizado pelas sociedades seguradoras.
```

```
#Através de análises estatísticas, foi verificado que a taxa de seguro, definida como a razão  
#entre prêmio médio e importância segurada média, observada para cada modelo de  
#veículo, #apresenta baixíssima variabilidade para as coberturas de Responsabilidade Civil  
#Danos #Materiais (RCDM), Responsabilidade Civil Danos Pessoais (RCDP), e Acidentes  
#Pessoais #de Passageiros (APP), quando comparada à variabilidade observada na
```

#cobertura Casco. #Por essa razão, foram incluídas na base para download as tabelas #PremReg e SinReg, #demonstrativas dos resultados classificados por região de circulação.

#A base possui também uma tabela que permite obter os dados de exposição, prêmios e #sinistros classificados por cidade de circulação.

#Descrição das tabelas:

#Tabelas Principais:

#arq_casco - contém dados de exposição, prêmios, sinistros e importância segurada para a #cobertura CASCO, classificados pela chave Categoria #Tarifária/Região/Modelo/Ano/Sexo/Faixa Etária;

#arq_casco3 - contém dados de exposição, prêmios e sinistros para a cobertura CASCO, #classificados pela chave Categoria Tarifária/CEP/Modelo/Ano;

#arq_casco4 - contém dados de exposição, prêmios e sinistros para a cobertura CASCO, #classificados pela chave Categoria Tarifária/Cidade/Modelo/Ano;

#premreg - distribuição regional de prêmios; e

#sinreg - distribuição regional de sinistros.

#Tabelas Auxiliares:

#auto2_vei - contém código FIPE e descrição de cada modelo de veículo, além do código do #grupo a que pertence;

#auto2_grupo - código e descrição dos grupos de modelos;

#auto_cat - código de descrição de categorias tarifárias;

#auto_cau - código e descrição de causas de sinistros;

#auto_CEP - correlaciona o CEP com cidades e regiões de circulação;

#auto_cob - código e descrição de coberturas;

#auto_idade - código e descrição de faixas etárias;

#auto_reg - código e descrição de regiões de circulação;

#auto_sexo - código e descrição de sexo (masculino, feminino, jurídico); e

#auto_cidade - código e nome das cidades.

###Histórico

#Até a versão disponibilizada em junho/2005, o sistema era fornecido em CD-ROM, de #forma que o usuário precisava instalar o programa e a base de dados completa em seu #computador.

#Os procedimentos necessários à instalação se revelaram demasiadamente complexos para #o usuário comum, o qual necessitava, eventualmente, consultar algumas informações. Em #consequência, o público-alvo do Sistema AUTOSEG praticamente se restringia às #companhias seguradoras, para subsidiar cálculos de tarifas.

#A implementação da versão 5.0, em 2006, teve por objetivo popularizar e ampliar o
#alcance #do Sistema, permitindo que as consultas sejam efetuadas diretamente pela
#Internet.

#A partir de 2008 a consulta on-line passou a abranger mais de um período e a base de
#dados #do último envio passou a ser disponibilizada no formato Access através do site da
#SUSEP

#Arquivo AutoSeg: auto_cau

#libraries

rm(list=ls(all=TRUE))

library(stringr)

library(dplyr)

#Dados#

#2016B

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2016B")

DadosASeg2016B=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2016B11Auxi = subset(DadosASeg2016B,DadosASeg2016B[,2]=="11")

str(DadosASeg2016B11Auxi)

EXPOSICAO <- scan(text=format(DadosASeg2016B11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2016B11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

CEP <- DadosASeg2016B11Auxi[,5]

DadosASeg2016B11Auxi2= DadosASeg2016B11Auxi[,-5]

DadosASeg2016B11Auxi3= cbind(DadosASeg2016B11Auxi2,[-5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2016B11Auxi3)

DadosASeg2016B11 <- DadosASeg2016B11Auxi3

SumASeg2016B11_ExpByCEP <-
aggregate(DadosASeg2016B11\$EXPOSICAO,by=list(DadosASeg2016B11\$CEP),
FUN=sum)

SumASeg2016B11_NSinRByCEP <- aggregate(DadosASeg2016B11\$FREQ_SIN1,
by=list(DadosASeg2016B11\$CEP), FUN=sum)

```

SumASeg2016B11_NSinPPPTByCEP <-
aggregate(DadosASeg2016B11$FREQ_SIN23,by=list(DadosASeg2016B11$CEP),
FUN=sum)

DadosASeg2016B11_ByCEP      <- cbind(SumASeg2016B11_ExpByCEP,
SumASeg2016B11_NSinRByCEP[,2], SumASeg2016B11_NSinPPPTByCEP[,2])

names(DadosASeg2016B11_ByCEP) <- c("CEP", "Exposicao",
"NSinRouboF","NSinPPPT")

head(DadosASeg2016B11_ByCEP)

sum(DadosASeg2016B11$EXPOSICAO)

write.csv2(DadosASeg2016B11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC II/Trabalho/bases/bases/DadosASeg2016B11_SumByCEP.csv")

#=====#

#2017A

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2017A")

DadosASeg2017A=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2017A11Auxi = subset(DadosASeg2017A,DadosASeg2017A[,2]=="11")

str(DadosASeg2017A11Auxi)

library(stringr)

CEP <- as.numeric(str_replace(DadosASeg2017A11Auxi[,5], "-", ""), digits=16)

EXPOSICAO <- scan(text=format(DadosASeg2017A11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2017A11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

DadosASeg2017A11Auxi2= DadosASeg2017A11Auxi[,-5]

DadosASeg2017A11Auxi3= cbind(DadosASeg2017A11Auxi2[,-5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2017A11Auxi3)

DadosASeg2017A11 <- DadosASeg2017A11Auxi3

SumASeg2017A11_ExpByCEP      <-
aggregate(DadosASeg2017A11$EXPOSICAO,by=list(DadosASeg2017A11$CEP),
FUN=sum)

SumASeg2017A11_NSinRByCEP    <- aggregate(DadosASeg2017A11$FREQ_SIN1,
by=list(DadosASeg2017A11$CEP), FUN=sum)

```

```

SumASeg2017A11_NSinPPPTByCEP <-
aggregate(DadosASeg2017A11$FREQ_SIN23,by=list(DadosASeg2017A11$CEP),
FUN=sum)

DadosASeg2017A11_ByCEP      <- cbind(SumASeg2017A11_ExpByCEP,
SumASeg2017A11_NSinRByCEP[,2], SumASeg2017A11_NSinPPPTByCEP[,2])

names(DadosASeg2017A11_ByCEP) <- c("CEP", "Exposicao",
"NSinRouboF","NSinPPPT")

head(DadosASeg2017A11_ByCEP)

sum(DadosASeg2017A11$EXPOSICAO)

write.csv2(DadosASeg2017A11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tu
do/UNIFESP/2020_1/TCC
II/Trabalho/bases/bases/DadosASeg2017A11_SumByCEP.csv")

#=====#

rm(list=ls(all=TRUE))

#2017B

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2017B")

DadosASeg2017B=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2017B11Auxi = subset(DadosASeg2017B,DadosASeg2017B[,2]=="11")

str(DadosASeg2017B11Auxi)

EXPOSICAO <- scan(text=format(DadosASeg2017B11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2017B11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

CEP <- DadosASeg2017B11Auxi[,5]

DadosASeg2017B11Auxi2= DadosASeg2017B11Auxi[,-5]

DadosASeg2017B11Auxi3= cbind(DadosASeg2017B11Auxi2[,-5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2017B11Auxi3)

DadosASeg2017B11 <- DadosASeg2017B11Auxi3

SumASeg2017B11_ExpByCEP      <-
aggregate(DadosASeg2017B11$EXPOSICAO,by=list(DadosASeg2017B11$CEP),
FUN=sum)

SumASeg2017B11_NSinRByCEP    <- aggregate(DadosASeg2017B11$FREQ_SIN1,
by=list(DadosASeg2017B11$CEP), FUN=sum)

```

```

SumASeg2017B11_NSinPPPTByCEP <-
aggregate(DadosASeg2017B11$FREQ_SIN23,by=list(DadosASeg2017B11$CEP),
FUN=sum)

DadosASeg2017B11_ByCEP      <- cbind(SumASeg2017B11_ExpByCEP,
SumASeg2017B11_NSinRByCEP[,2], SumASeg2017B11_NSinPPPTByCEP[,2])

names(DadosASeg2017B11_ByCEP) <- c("CEP", "Exposicao",
"NSinRouboF","NSinPPPT")

head(DadosASeg2017B11_ByCEP)

sum(DadosASeg2017B11$EXPOSICAO)

write.csv2(DadosASeg2017B11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC II/Trabalho/bases/bases/DadosASeg2017B11_SumByCEP.csv")

#=====#

rm(list=ls(all=TRUE))

#2018A

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2018A")

DadosASeg2018A=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2018A11Auxi = subset(DadosASeg2018A,DadosASeg2018A[,2]=="11")

str(DadosASeg2018A11Auxi)

EXPOSICAO <- scan(text=format(DadosASeg2018A11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2018A11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

CEP <- DadosASeg2018A11Auxi[,5]

DadosASeg2018A11Auxi2= DadosASeg2018A11Auxi[,-5]

DadosASeg2018A11Auxi3= cbind(DadosASeg2018A11Auxi2[,5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2018A11Auxi3)

DadosASeg2018A11 <- DadosASeg2018A11Auxi3

SumASeg2018A11_ExpByCEP      <-
aggregate(DadosASeg2018A11$EXPOSICAO,by=list(DadosASeg2018A11$CEP),
FUN=sum)

```

```

SumASeg2018A11_NSinRByCEP <- aggregate(DadosASeg2018A11$FREQ_SIN1,
by=list(DadosASeg2018A11$CEP), FUN=sum)

SumASeg2018A11_NSinPPPTByCEP <-
aggregate(DadosASeg2018A11$FREQ_SIN23,by=list(DadosASeg2018A11$CEP),
FUN=sum)

DadosASeg2018A11_ByCEP <- cbind(SumASeg2018A11_ExpByCEP,
SumASeg2018A11_NSinRByCEP[,2], SumASeg2018A11_NSinPPPTByCEP[,2])

names(DadosASeg2018A11_ByCEP) <- c("CEP", "Exposicao",
"NSinRouboF", "NSinPPPT")

head(DadosASeg2018A11_ByCEP)

sum(DadosASeg2018A11$EXPOSICAO)

write.csv2(DadosASeg2018A11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tu
do/UNIFESP/2020_1/TCC
II/Trabalho/bases/bases/DadosASeg2018A11_SumByCEP.csv")

#=====#

rm(list=ls(all=TRUE))

#2018B

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2018B")

DadosASeg2018B=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2018B11Auxi = subset(DadosASeg2018B,DadosASeg2018B[,2]=="11")

str(DadosASeg2018B11Auxi)

EXPOSICAO <- scan(text=format(DadosASeg2018B11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2018B11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

CEP <- DadosASeg2018B11Auxi[,5]

DadosASeg2018B11Auxi2= DadosASeg2018B11Auxi[,-5]

DadosASeg2018B11Auxi3= cbind(DadosASeg2018B11Auxi2[,-5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2018B11Auxi3)

DadosASeg2018B11 <- DadosASeg2018B11Auxi3

SumASeg2018B11_ExpByCEP <-
aggregate(DadosASeg2018B11$EXPOSICAO,by=list(DadosASeg2018B11$CEP),
FUN=sum)

```

```

SumASeg2018B11_NSinRByCEP <- aggregate(DadosASeg2018B11$FREQ_SIN1,
by=list(DadosASeg2018B11$CEP), FUN=sum)

SumASeg2018B11_NSinPPPTByCEP <-
aggregate(DadosASeg2018B11$FREQ_SIN23,by=list(DadosASeg2018B11$CEP),
FUN=sum)

DadosASeg2018B11_ByCEP <- cbind(SumASeg2018B11_ExpByCEP,
SumASeg2018B11_NSinRByCEP[,2], SumASeg2018B11_NSinPPPTByCEP[,2])

names(DadosASeg2018B11_ByCEP) <- c("CEP", "Exposicao",
"NSinRouboF", "NSinPPPT")

head(DadosASeg2018B11_ByCEP)

sum(DadosASeg2018B11$EXPOSICAO)

write.csv2(DadosASeg2018B11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC II/Trabalho/bases/bases/DadosASeg2018B11_SumByCEP.csv")

#=====#

#2019A

setwd("C:/Users/JB/Desktop/Maio2019/OUTROS2/Tudo/UNIFESP/2020_1/TCC
II/Trabalho/bases/Autoseg2019A")

DadosASeg2019A=read.csv("arq_casco3_comp.csv",sep = ";", header = TRUE)

DadosASeg2019A11Auxi = subset(DadosASeg2019A,DadosASeg2019A[,2]=="11")

str(DadosASeg2019A11Auxi)

library(stringr)

CEP <- as.numeric(str_replace(DadosASeg2019A11Auxi[,5], "-", ""), digits=16)

EXPOSICAO <- scan(text=format(DadosASeg2019A11Auxi[,6], decimal.mark=","),
dec="," , sep=".")

PREMIO<- scan(text=format(DadosASeg2019A11Auxi[,7], decimal.mark=","), dec="," ,
sep=".")

DadosASeg2019A11Auxi2= DadosASeg2019A11Auxi[,-5]

DadosASeg2019A11Auxi3= cbind(DadosASeg2019A11Auxi2[,-5:-
6],CEP,EXPOSICAO,PREMIO)

str(DadosASeg2019A11Auxi3)

DadosASeg2019A11 <- DadosASeg2019A11Auxi3

SumASeg2019A11_ExpByCEP <-
aggregate(DadosASeg2019A11$EXPOSICAO,by=list(DadosASeg2019A11$CEP),
FUN=sum)

SumASeg2019A11_NSinRByCEP <- aggregate(DadosASeg2019A11$FREQ_SIN1,
by=list(DadosASeg2019A11$CEP), FUN=sum)

```



```
SumASeg2019A11_NSinPPPTByCEP <-  
aggregate(DadosASeg2019A11$FREQ_SIN23,by=list(DadosASeg2019A11$CEP),  
FUN=sum)  
  
DadosASeg2019A11_ByCEP      <- cbind(SumASeg2019A11_ExpByCEP,  
SumASeg2019A11_NSinRByCEP[,2], SumASeg2019A11_NSinPPPTByCEP[,2])  
  
names(DadosASeg2019A11_ByCEP) <- c("CEP", "Exposicao",  
"NSinRouboF", "NSinPPPT")  
  
head(DadosASeg2019A11_ByCEP)  
  
sum(DadosASeg2019A11$EXPOSICAO)  
  
write.csv2(DadosASeg2019A11_ByCEP,"C:/Users/JB/Desktop/Maio2019/OUTROS2/Tu  
do/UNIFESP/2020_1/TCC  
II/Trabalho/bases/bases/DadosASeg2019A11_SumByCEP.csv")
```